ACHI 2013

The Sixth International Conference on Advances in Computer-Human Interactions

ISBN: 978-1-61208-250-9

February 24 - March 1, 2013

Nice, France

ACHI 2013 Editors

Leslie Miller, Iowa State University - Ames, USA

ACHI 2013

Forward

The sixth edition of The International Conference on Advances in Computer-Human Interactions (ACHI 2013) conference was held in Nice, France, February 24 - March 1, 2013.

The conference on Advances in Computer-Human Interaction, ACHI 2013, was a result of a paradigm shift in the most recent achievements and future trends in human interactions with increasingly complex systems. Adaptive and knowledge-based user interfaces, universal accessibility, human-robot interaction, agent-driven human computer interaction, and sharable mobile devices are a few of these trends. ACHI 2013 brought also a suite of specific domain applications, such as gaming, social, medicine, education and engineering.

The event was very competitive in its selection process and very well perceived by the international scientific and industrial communities. As such, it is attracting excellent contributions and active participation from all over the world. We were very pleased to receive a large amount of top quality contributions.

The accepted papers covered a wide range of human-computer interaction related topics such as graphical user interfaces, input methods, training, recognition, and applications.
We believe that the ACHI 2013 contributions offered a large panel of solutions to key problems in all areas of human-computer interaction.

We take here the opportunity to warmly thank all the members of the ACHI 2013 technical program committee as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and efforts to contribute to the ACHI 2013. We truly believe that thanks to all these efforts, the final conference program consists of top quality contributions.

This event could also not have been a reality without the support of many individuals, organizations and sponsors. In addition, we also gratefully thank the members of the ACHI 2013 organizing committee for their help in handling the logistics and for their work that is making this professional meeting a success.

We hope the ACHI 2013 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the human-computer interaction field.

We also hope that Côte d’Azur provided a pleasant environment during the conference and everyone saved some time for exploring the Mediterranean Coast.

ACHI 2013 Chairs

Leslie Miller, Iowa State University - Ames, USA
Silvana Roncagliolo, Pontificia Universidad Católica de Valparaíso, Chile
Raymond Jarvis, Monash University, Australia
Katia Sykara, Carnegie Mellon University, USA

ACHI 2013

Committee

ACHI Advisory Committee

Leslie Miller, Iowa State University - Ames, USA
Silvana Roncagliolo, Pontificia Universidad Católica de Valparaíso, Chile
Raymond Jarvis, Monash University, Australia
Katia Sykara, Carnegie Mellon University, USA

ACHI 2013 Technical Program Committee

Andrea Abate, University of Salerno, Italy
Víctor Acinas Garzón, Instalaciones Inabensa S.A. - Madrid, Spain
Eugenio Aguirre Molina, University of Granada, Spain
Margherita Antona, Foundation for Research and Technology - Hellas (FORTH) - Heraklion, Greece
Marios Belk, University of Cyprus, Cyprus
Pierre Borne, Ecole Centrale de Lille / IEEE Fellow, France
Marius Brade, Technische Universität Dresden, Germany
Luis Carriço, University of Lisbon, Portugal
Jessie Y.C. Chen, U.S. Army Research Laboratory - Human Research & Engineering Directorate, USA
Ping Chen, University of Houston-Downtown, USA
Jacek Chmielewski, Poznań University of Economics, Poland
Ryszard S. Choras, University of Technology & Life Sciences - Bydgoszcz, Poland
Shaowei Chu, University of Tsukuba, Japan
José Coelho, University of Lisbon, Portugal
Cesar Alberto Collazos O., University of Cauca - Popayán, Colombia
Dragos Datcu, Netherlands Defence Academy, The Netherlands
Kerstin Dautenhahn, The University of Hertfordshire, United Kingdom
Lucio Tommaso De Paolis, Salento University - Lecce, Italy
Nancy Diniz, Xi'an Jiaotong-Liverpool University, China
John M. Dolan, Carnegie Mellon University, USA
Rolf Drechsler, DFKI - Bremen, Germany
Carlos Duarte, University of Lisbon, Portugal
Martin Ebner, Graz University of Technology, Austria
Anastasios A. Economides University of Macedonia - Thessaloniki, Hellas
Larbi Esmahi, Athabasca University, Canada
Elisabetta Farella, DEIS - University of Bologna, Italy
Michael Farrugia, University College Dublin, Ireland
Karim Fraoua, Université Paris-Est- Conservatoire des arts et métiers, France
Somchart Fugkeaw, Mahidol University - Nakhonpathom, Thailand
Mathias Funk, Eindhoven University of Technology, The Netherland
Miguel García, Universidad Politecnica de Valencia, Spain
Matjaz Gams, Josef Stefan Institute- Ljubljana, Slovenia

Marta Gatius, Universitat Politècnica de Catalunya, Spain
Cristina Gena, Università di Torino, Italy
Denis Gillet, Swiss Federal Institute of Technology (EPFL), Switzerland
Mehmet Göktürk, Gebze Institute of Technology, Turkey
Luis Gomes, Universidade Nova de Lisboa / UNINOVA-CTS - Monte de Caparica, Portugal
Pascual Gonzalez, University of Castilla-La Mancha, Spain
Bernard Grabot, LGP-ENIT, France
Andrina Granic, University of Split, Croatia
Vic Grout, Glyndwr University - Wrexham,UK
Jaume Gual Orti, Universitat Jaume I – Castellón, Spain
Ido Guy, IBM Research-Haifa, Israel
Maki K. Habib, The American University in Cairo, Egypt
Till Røssvoll Halbach, Norwegian Computing Center / Norsk Regnesentral (NR), Norway
Kristian J. Hammond, Northwestern University - Evanston, USA
Christoph Hintermüller, Guger Technologies OG – Schiedlberg, Austria
Jochen Hirth, University of Kaiserslautern, Germany
Thomas Holz, University College Dublin, Ireland
Weidong (Tony) Huang, CSIRO ICT Centre, Australia
Ilias Iliadis, IBM Research - Zurich, Switzerland
Jamshed Iqbal, Italian Institute of Technology (IIT) - Genova, Italy
Sandy Ingram-El Helou, EPFL - Lausanne, Switzerland
Dragan Ivetic, University of Novi Sad, Serbia
Raymond Jarvis, Monash University, Australia
Ran Jiao, Beihang University - Beijing, China
Gareth Jones, Dublin City University, Ireland
Hanmin Jung, Korea Institute of Science & Technology Information, Korea
Janusz Kacprzyk, Polish Academy of Sciences - Warsaw / IEEE Fellow, Poland
Hermann Kaindl, Vienna University of Technology, Austria
Ahmed Kamel, Concordia College - Moorhead, USA
Athanasios Kapoutsis, Centre for Research and Technology Hellas | Information Technologies Institute, Greece
Koji Kashihara, Tokushima University, Japan
Etienne Kerre, Retired Professor, Belgium
Hamed Ketabdar, Deutsche Telekom Laboratories - TU Berlin, Germany
Jonghwa Kim, University of Augsburg, Germany
Michael Kipp, University of Applied Sciences Augsburg, Germany
Frank Klawonn, Ostfalia University of Applied Sciences - Wolfenbuettel, Germany
Anita Komlodi, University of Maryland - Baltimore County, USA
Satoshi Kurihara, Osaka University, Japan
Andrew Kusiak, The University of Iowa, USA
Vidas Lauruska, Siauliai University, Lithuania
Cheongjae Lee, Kyoto University, Japan
Soo-Young Lee, KAIST and RIKEN BSI, Korea
Alma Leora Culén , University of Oslo, Norway
Cynthia Y. Lester, Tuskegee University, USA
Tsai-Yen Li, National Chengchi University - Taipei, Taiwan
Yangmin Li, University of Macau - Taipa, Macao
Chin-Teng (CT) Lin, National Chiao-Tung University, Taiwan

Jean-Luc Lugrin, Universität Würzburg, Germany
Ameesing Luximon, Hong Kong Polytechnic University, Hong Kong
Francisco Madera, Universidad Autonoma de Yucatan - Merida, Mexico
Jacek Mandziuk, Warsaw University of Technology, Poland
Antonio Maria Rinaldi, Università di Napoli Federico II, Italy
Christoph Mayer, Technische Universität München, Germany
Gerrit Meixner, German Research Center for Artificial Intelligence (DFKI) - Kaiserslautern, Germany
Harald Milchrahm, Technical University Graz, Austria
Leslie Miller, Iowa State University - Ames, USA
Yasser F. O. Mohammad, Assiut University, Egypt
Jaime Muñoz-Arteaga, Universidad Autónoma de Aguascalientes, México
Fazel Naghdy, University of Wollongong, Australia
Toru Nakata, National Institute of Advanced Science and Technology (AIST), Japan
Deok Hee Nam, Wilberforce University, USA
Andreas C. Nearchou, University of Patras, Greece
Roger Ng, Hong Kong Polytechnic University, Hong Kong
Radoslaw Niewiadomski , Telecom ParisTech, France
Gregory O'Hare, University College Dublin (UCD), Ireland
Fakri Othman, Cardiff Metropolitan University (UWIC), UK
Volker Paelke, Institut de Geomàtica – Castelldefels/Barcelona, Spain
Sethuraman (Panch) Panchanathan, Arizona State University - Tempe, USA
Evangelos Papadopoulos, National Technical University of Athens, Greece
Igor Paromtchik, INRIA - Saint Ismier, France
Alexandru Patriciu. McMaster University - Hamilton, Canada
Estelle Perrin, CReSTIC | Reims University – Charleville-Mezieres, France
Igor Petukhov, Blitzkrieg, Russia
Jose Pinto, Universidade do Porto, Portugal
Marina Puyuelo Cazorla, Universitat Politècnica de València (UPV), Spain
Didier Puzenat, Université des Antilles et de la Guyane - Pointe-à-Pitre, Guadeloupe
Rafael Radkowski, Iowa State University, USA
Nitendra Rajput, IBM Research - New Delhi, India
Xiangshi Ren, Kochi University of Technology, Japan
Andreas Riener, University Linz - Linz/Donau, Austria
Joel Rodrigues, Intituto de Telecomunicações / University of Beira Interior, Portugal
Silvana Roncagliolo, Pontificia Universidad Católica de Valparaíso, Chile
Agostinho Rosa, Lasseb-ISR-IST / Technical University of Lisbon, Portugal
José Rouillard, Université des Sciences et Technologies de Lille, France
Rebekah Rousi, University of Jyväskylä, Finland
Juha Röning, University of Oulu, Finland
Antonio José Sánchez Salmerón, Universidad Politécnica de Valencia, Spain
Munehiko Sasajima, Osaka University, Japan
Dominique Scapin, INRIA - Le Chesnay, France
Marc Seißler, German Research Center for Artificial Intelligence (DFKI) - Kaiserslautern, Germany
Shunji Shimizu, Tokyo University of Science, Suwa, Japan
Pavel Slavik, Czech Technical University in Prague, Czech Republic
Ross Smith, University of South Australia, Australia
Jaka Sodnik, University of Ljubljana, Slovenia
Weilian Su, Naval Postgraduate School - Monterey, USA

Copyright Information

Table of Contents

Investigating Players' Affective States in an Interactive Environment *Uttam Kokil*	1
Emergent Design System Using Computer-Human Interactions and Serendipity *Akira Kito, Yuki Mizumachi, Koichiro Sato, and Yoshiyuki Matsuoka*	7
Sensory Evaluation Method to Create Pictograms Based on Multiplex Sign Languages *Naotsune Hosono, Hiromitsu Inoue, Yuji Nagashima, and Yutaka Tomita*	13
Accessibility and Augmented Reality into Heritage Site Interpretation - A Pilot Experience with Visitors at the Monument Lonja de la Seda in Valencia (Spain) *Marina Puyuelo Cazorla, Jose Luis Higon Calvet, Monica Val Fiel, and Lola Merino Sanjuan*	17
A Deported View Concept for Touch Interaction *Alexandre Alapetite, Rune Fogh, Henning Boje Andersen, and Ali Gurcan Ozkil*	22
Effect of non-Unified Interaction Design of in-car Applications on Driving Performance, Situational Awareness and Task Performance *Julia Manner, Christopher Kohl, Michael Schermann, and Helmut Krcmar*	28
Studying Depth in a 3D User Interface by a Paper Prototype as a Part of the Mixed Methods Evaluation Procedure. Early Phase User Experience Study *Leena Arhippainen, Minna Pakanen, and Seamus Hickey*	35
Studying Four 3D GUI Metaphors in Virtual Environment in Tablet Context. Visual Design and Early Phase User Experience Evaluation *Minna Pakanen, Leena Arhippainen, and Seamus Hickey*	41
Towards a 3D User Interface in a Tablet Device Context. An Iterative Design and Evaluation Process *Leena Arhippainen, Minna Pakanen, and Seamus Hickey*	47
Subjective Usability of Speech, Touch and Gesture in a Heterogeneous Multi-Display Environment *Arnoud P.J. de Jong, Susanne Tak, Alexander Toet, Sven Schultz, Jan Pieter Wijbenga, and Jan van Erp*	53
Dynamic Gesture Recognition Based on Fuzzy Neural Network Classifier *Ching-Han Chen, Kirk Chang, Nai-Yuan Liu, and Gimmy Su*	57
BrainBrush, a Multimodal Application for Creative Expressivity *Bram van de Laar, Ivo Brugman, Femke Nijboer, Mannes Poel, and Anton Nijholt*	62
Multimodal Human-Robot Interactions: the Neurorehabilitation of Severe Autistic Children	68

Irini Giannopulu

What Should a Robot do for you? - Evaluating the Needs of the Elderly in the UK 83
Hagen Lehmann, Dag Sverre Syrdal, Kerstin Dautenhahn, GertJan Gelderblom, Sandra Bedaf, and Farshid Amirabdollahian

Investigating Child-Robot Tactile Interactions: A Taxonomical Classification of Tactile Behaviour of Children with Autism Towards a Humanoid Robot. 89
Ben Robins, Farshid Amirabdollahian, and Kerstin Dautenhahn

Resource-Efficient Methods for Feasibility Studies of Scenarios for Long-Term HRI Studies 95
Nate Derbinsky, Wan Ching Ho, Ismael Duque, Joe Saunders, and Kerstin Dautenhahn

Person Identification using Skeleton Information from Kinect 101
Aniruddha Sinha, Kingshuk Chakravarty, and Brojeshwar Bhowmick

Robot Learning Rules of Games by Extraction of Intrinsic Properties 109
Gregoire Pointeau, Maxime Petit, and Peter Ford Dominey

”Where is Your Nose?” - Developing Body Awareness Skills Among Children With Autism Using a Humanoid Robot 117
Sandra Costa, Hagen Lehmann, Ben Robins, Kerstin Dautenhahn, and Filomena Soares

An Interactive Game with a Robot: Peoples' Perceptions of Robot Faces and a Gesture-Based User Interface 123
Michael Walters, Samuel Marcos, Dag Sverre Syrdal, and Kerstin Dautenhahn

Robust Perception of an Interaction Partner Using Depth Information 129
Salah Saleh, Anne Kickton, Jochen Hirth, and Karsten Berns

Gesture Recognition for Humanoid Assisted Interactive Sign Language Tutoring 135
Bekir Sitki Ertugrul, Cemal Gurpinar, Hasan Kivrak, Ajla Kulaglic, and Hatice Kose

Knowledge-driven User Activity Recognition for a Smart House. Development and Validation of a Generic and Low-Cost, Resource-Efficient System 141
Ismael Duque, Kerstin Dautenhahn, Kheng Lee Koay, Ian Willcock, and Bruce Christianson

Comparison of Simultaneous Measurement of Lens Accommodation and Convergence in Viewing Natural and Stereoscopic Visual Target 147
Tomoki Shiomi, Takehito Kojima, Keita Uemoto, and Masaru Miyao

Study of a FCMAC ANN for Implementation in the Modeling of an Active Control Transtibial Prosthesis 151
Jose Alberto Alves Andrade, Lourdes Mattos Brasil, Everaldo Henrique Diniz, Keli Cristina Vieira Siqueira Borges, Jeann Feitosa Figueiredo, and Rita de Cassia Silva

The Iterative Design and Evaluation Approach for a Socially-aware Search and Retrieval Application for Digital Archiving 157
Dimitris Spiliotopoulos, Ruben Bouwmeester, Dominik Frey, Georgios Kouroupetroglou, and Pepi Stavropoulou

Effect of Agent Embodiment on the Elder User Enjoyment of a Game 162
Jeremy Wrobel, Ya Huei Wu, Helene Kerherve, Laila Kamali, Anne Sophie Rigaud, Celine Jost, Brigitte Le Pevedic, and Dominique Duhaut

Augmenting Remote Trading Card Play with Virtual Characters used in Animation and Game Stories - Towards Persuasive and Ambient Transmedia Storytelling - 168
Mizuki Sakamoto, Todorka Alexandrova, and Tatsuo Nakajima

An Interactive Agent Supporting First Meet Based on Adaptive Entrainment Control 178
Tatsuya Hayamizu, Mutsuo Sano, Kenzaburo Miyawaki, and Kentarou Mukai

The Virtual Counselor - Automated Character Animation for Ambient Assisted Living 184
Sascha Fagel, Martin Morandell, Christopher Mayer, and Andreas Hilbert

AlgoPath's New Interface Helps You Find Your Way Through Common Algorithmic Mistakes 188
Estelle Perrin and Sebastien Linck

A Three-Dimensional Interactive Simulated-Globe System Application in Education 194
Wei-Kai Liou and Chun-Yen Chang

Virtual Simulation of the Construction Activity: Bridge Deck Composed of Precast Beams 199
Luis Viana and Alcinia Zita Sampaio

The iPad in a Classroom: A Cool Personal Item or Simply an Educational Tool? 204
Andrea Gasparini and Alma Leora Culen

Architecture of an Intelligent Tutoring System Applied to the Breast Cancer Based on Ontology, Artificial Neural Networks and Expert Systems 210
Henrique P. Maffon, Jairo S. Melo, Tamara A. Morais, Patrycia B. Klavdianos, Lourdes M. Brasil, Thiago L. Amaral, and Gloria M. Curilem

CyPhy-UI: Cyber-Physical User Interaction Paradigm to Control Networked Appliances with Augmented Reality 215
Kenya Sato, Naoya Sakamoto, Shinya Mihara, and Hideki Shimada

Luminance Contrast Influences Reaction Time in Young and Older Adults 222
Patrick J. Grabowski and Andrea H. Mason

Networked Visibility: The case of smart card ticket information 228
Maja van der Velden, Alma Leora Culen, Jo Herstad, and Abdulhakeem Atif

TV Applications for the Elderly: Assessing the Acceptance of Adaptation and Multimodality 234
Jose Coelho, Pradipta Biswas, Tiago Guerreiro, Gokcen Aslan, Carlos Duarte, and Pat Langdon

Identifying Cross-Platform and Cross-Modality Interaction Problems in e-Learning Environments 243
Andre da Silva, Fernanda Freire, and Heloisa da Rocha

Applying Commercial Digital Games to Promote Upper Extremity Movement Functions for Stroke Patients 250
Lan-Ling Huang, Chang-Franw Lee, and Mei-Hsiang Chen

Evaluating the Interaction of Users with Low Vision in a Multimodal Environment 256
Clodis Boscarioli, Marcio Seiji Oyamada, Jorge Bidarra, and Marcelo Fudo Rech

Bimanual Performance in Unpredictable Virtual Environments: A Lifespan Study 263
Andrea Mason, Drew Rutherford, and Patrick Grabowski

Usability Analysis of Children's iPad Electronic Picture books 269
Pei-shiuan Tsai and Manlai You

Evaluating the Impact of Spatial Ability in Virtual and Real World Environments 274
Georgi Batinov, Kofi Whitney, Les Miller, Sarah Nusser, Bryan Stanfill, and Kathleen Ashenfelter

Software Lifecycle Activities to Improve Security Into Medical Device Applications 280
Diogo Rispoli, Vinicius Rispoli, Paula Fernandes, and Lourdes Brasil

“Handreha”: A new Hand and Wrist Haptic Device for Hemiplegic Children 286
Mohamed Bouri, Charles Baur, Reymond Clavel, Christopher John Newman, and Milan Zedka

Fundamental Study to Consider for Evaluation of A Welfare Device 293
Hiroaki Inoue, Shunji Shimizu, Noboru Takahashi, Hiroyuki Nara, Takeshi Tsuruga, Fumikazu Miwakeichi, Nobuhide Hirai, Senichiro Kikuchi, Eiju Watanabe, and Satoshi Kato

Haptic Manipulation of Objects on Multitouch Screens: Effects of Screen Elevation, Inclination and Task Requirements on Posture and Fatigue 299
Samantha Scotland, Shwetarupalika Das, Thomas Armstrong, and Bernard Martin

Sliding Raised-Dots Perceptual Characteristics 303
Yoshihiko Nomura, Syed Muammar Najib Syed Yusoh, Kazuki Iwabu, and Ryota Sakamoto

1 DOF Tabletop Haptic Mouse for Shape Recognition of 3D Virtual Objects 309
Hiroshi Suzuki, Hiroaki Yano, and Hiroo Iwata

Usability Study of Static/Dynamic Gestures and Haptic Input as Interfaces to 3D Games 315
Farzin Farhadi-Niaki, Jesse Gerroir, Ali Arya, S. Ali Etemad, Robert Laganiere, Pierre Payeur, and Robert

Biddle

Interactive Dynamic Simulations with Co-Located Maglev Haptic and 3D Graphic Display 324
Peter Berkelman, Sebastian Bozlee, and Muneaki Miyasaka

Haptic System for Eyes Free and Hands Free Pedestrian Navigation 330
Nehla Ghouaiel, Jean-Marc Cieutat, and Jean-Pierre Jessel

Haptic Mouse - Enabling Near Surface Haptics in Pointing Interfaces 336
Kasun Karunanayaka, Sanath Siriwardana, Chamari Edirisinghe, Ryohei Nakatsu, and Ponnampalam Gopalakrishnakone

Fundamental Study to Consider for Advanced Interface in Grasping Movement 342
Shunji Shimizu, Hiroaki Inoue, and Noboru Takahashi

Analyzing the Effects of Virtualizing and Augmenting Trading Card Game based on the Player's Personality 348
Mizuki Sakamoto, Todorka Alexandrova, and Tatsuo Nakajima

Influence of Relationship between Game Player and Remote Player on Emotion 358
Masashi Okubo, Tsubasa Yamashita, and Mamiko Sakata

Reducing the User Burden of Identity Management: A Prototype Based Case Study for a Social-Media Payment Application 364
Till Halbach Rossvoll and Lothar Fritsch

Heads Up: Using Cognitive Mapping to Develop a Baseline Description for Urban Visualization 371
Ginette Wessel, Elizabeth Unruh, and Eric Sauda

User Support System for Designing Decisional Database 377
Fatma Abdelhedi and Gilles Zurfluh

Information Needs Of Chinese Mobile Internet Users: A User Study 383
Yanxia Yang and Grace Deng

Emotion Recognition using Autonomic Nervous System Responses 389
Byoung-Jun Park, Eun-Hye Jang, Sang-Hyeob Kim, Chul Huh, Myoung-Ae Chung, and Jin-Hun Sohn

Classification of Human Emotions from Physiological signals using Machine Learning Algorithms 395
Eun-Hye Jang, Byoung-Jun Park, Sang-Hyeob Kim, Myoung-Ae Chung, Mi-Sook Park, and Jin-Hun Sohn

Filling the User Skill Gap Using HCI Techniques to Implement Experimental Protocol on Driving Simulators 401
Ghasan Bhatti, Roland Bremond, Jean-Pierre Jessel, Nguyen-thong Dang, Fabrice Vienne, and Guillaume Millet

Instrumentation and Features Selection Using a Realistic Car Simulator in Order to Perform Efficient Single-User Drunkenness Analysis 407
Audrey Robinel and Didier Puzenat

Adaptive Simulation of Monitoring Behavior: The Adaptive Information Expectancy Model 413
Bertram Wortelen and Andreas Ludtke

Shape Modeling: From Linear Anthropometry to Surface Model 420
Ameersing Luximon and Huang Chao

A Modular Interface Design to Indicate a Robot's Social Capabilities 426
Frank Hegel

Automatic Discrimination of Voluntary and Spontaneous Eyeblinks. Use of the blink as a switch interface 433
Shogo Matsuno, Minoru Ohyama, Kiyohiko Abe, Hironobu Sato, and Shoichi Ohi

Cursor Control Trace: Another look into eye-gaze, hand, and eye-hand pointing techniques 440
Ricardo Sol, Mon-Chu Chen, and Jose Carlos Marques

A Hybrid Tracking Solution to Enhance Natural Interaction in Marker-based Augmented Reality Applications 444
Rafael Radkowski and James Oliver

Evaluating Multi-Modal Eye Gaze Interaction for Moving Object Selection 454
Jutta Hild, Elke Muller, Edmund Klaus, Elisabeth Peinsipp-Byma, and Jurgen Beyerer

Touch-Screens and Elderly users: A Perfect Match? 460
Alma Leora Culen and Tone Bratteteig

Basic Study for New Assistive System Based on Brain Activity during Car Driving 466
Shunji Shimizu, Hiroaki Inoue, Hiroyuki Nara, Noboru Takahashi, Fumikazu Miwakeichi, Nobuhide Hirai, Senichiro Kikuchi, Eiju Watanabe, and Satoshi Kato

Proposal of an Automobile Driving Interface Using Gesture Operation for Disabled People 472
Yoshitoshi Murata, Kazuhiro Yoshida, Kazuhiro Suzuki, and Daisuke Takahashi

The European MobileSage Project – Situated Adaptive Guidance for the Mobile Elderly 479
Till Halbach Rossvoll

e-Learning Environment with Multimodal Interaction: A proposal to improve the usability, accessibility and learnability of e-learning environments 483
Andre da Silva and Heloisa da Rocha

A.M.B.E.R. Shark-Fin: An Unobtrusive Affective Mouse 488

Thomas Christy and Ludmila I. Kuncheva

Investigating Players' Affective States in an Interactive Environment

Uttam Kokil
Department of Technology, CoBPA
UND, Grand Forks, ND 58202. USA
Email: uttam.kokil@und.edu

***Abstract*—The objective of this paper is to examine player experience from a Human Computer Interaction Design perspective whereby usability, aesthetics, and hedonic components can be investigated in an interactive domain. Mihaly Csikszentmihalyi's concept of flow has been applied to the Component of User Experience (CUE) model to measure user experience (UX) of products, and other instruments such as the Presence Involvement Flow Framework *(ver. 2)* (PIFF); Gameflow and Game Experience Questionnaire to evaluate game enjoyment respectively. So far, the CUE Model has been applied in simulated user-testing situation. It becomes important to gauge the potential of the CUE Model in an interactive game scenario, given that it is composed of components such as perceived usability, perceived aesthetics and emotional responses. The aim of this study is to conduct a comparative analysis of two user experience models (Component of User Experience) and PIFF to examine players' emotional responses in four different conditions (low and high usability, low and high aesthetic value) when they play two different computer game category, namely "Hard Fun" and "Easy Fun" game respectively, for the touchscreen. This research will investigate how two independent variables, usability (low and high) and visual aesthetic (low and high), of a game user interface design will affect the dependent variables: player experience, task performance and emotional responses (enjoyment).**

Keywords-user experience; visual aesthetics; computer games; game usability.

I. INTRODUCTION

Researchers argue that experiences are the new area of economic growth among consumers [1]. The game designer creates various experiences in a game in order to generate player enjoyment. *PIFF* [2] is a validated questionnaire that can be used to investigate user experience in games. The *Component of User Experience (CUE) Model* describes the three fundamental components of user experience that is comprised of instrumental quality, non-instrumental quality and emotional responses [3]. *CUE Model* has been applied to simulated product scenarios (e.g. simulations of audio players) however we are interested to examine its potential in an interactive environment to understand how it can be appropriated into the domain of computer game. It is important to investigate if *CUE Model* shares the same validity as another user experience tool such as *PIFF*, and whether it can be utilized to study "*users' experience of interaction with interactive products and systems*," in a gaming environment, as is the essence of this model.

Video and computer games are the fastest growing form of entertainment. The revenue generated by the computer game industry escalated to US$ 16.6 billion in 2011 from US$ 5.5 billion in 1999, in the U.S.A [4]. Technology is evolving at such an unprecedented pace that many end-users suddenly find themselves confronted with new mode of interaction especially when it comes to computer gaming with multi-touch screen or other controls. User experience varies in different environments [5], on different display sizes [6] or when using different input devices. Technology alone does not make a game successful. Player experience forms part and parcel of the gaming process. For instance, despite the rise of gaming technology, players have continued to express dissatisfaction with current computer games because most games do not provide the optimal level of challenge to accommodate players' skills, learning capabilities, and emotional qualities [7,8]. It is said that certain games are either too easy to win, or that they do not challenge the player to use his or her skills appropriately, and as a consequence, the player does not derive enough enjoyment during game play [7]. Lazarro explains that only 20% of the games that hit the market are successful [10]. It becomes imperative to investigate the phenomena that create optimum player experience in computer games by exploring the components of player experience as regards to the graphical user interface. Kankainen [11] defines user experience as a "*result of a motivated action in a certain context. User's previous experiences and expectations influence the present experience; this present experience leads to more experiences and modified expectations."*

Until recently, more attention was given solely to the usability of a product, while other aspects that formed part of user experience were neglected. Gaming has received more attention in the HCI area, but "there have been only a few attempts to study user experience in games holistically." [5]. There are several components in a user interface design that give rise to different types of cognitive and emotional responses among users during interaction. Interaction design creates "user experiences that enhance and extend the way people work, communicate and interact" [12]. "User experience is highly subjective and dependent on a user's expectations, which in turn are shaped by user's age, education, cultural background, gender and several other factors" [13]. Researchers recognize the fact that games evoke an intensified level of emotions during game play so as to make the experience entertaining and enjoyable [14].

 ISBN: 978-1-61208-250-9

Aesthetics have been found to affect users' perceptions toward product use [15]. Lindgaard and Dudek [16] state that aesthetics is directly associated to emotions through the senses. Don Norman [17] also mentions that appearance of a product affects the very visceral part.

This research is designed to study the theory of the "four Fun Keys." Lazarro [18] describes game classification in terms of Fun Keys: "Hard Fun, Easy Fun, Altered state and People factor." This theory explains how each game category provides the necessary motivation to play, and the spirits of the gamer in each type of game. The player is determined to meet specific goals when playing a hard fun game that elicits emotions such as frustration and "fiero." For example, the scores achieved gauge the skill level. In an Easy Fun type game, the players explore and discover the game by "focusing on the enjoyment of experiencing the game activities." This is not about winning the game but rather it focuses and stimulates one's curiosity by figuring out the different game strategies and levels. In the Altered state, players feel an internal change, from one mental state to another. People factor is more about the experience a player derives while interacting with other players during gameplay such as in multiplayer games.

For this research, "visceral rhetoric" can be defined as the emotional responses that are automatically triggered by audio and visual components of the game surface layer. Game design comprises of game interface, game mechanics and Game Play [19]. This indicates that game interface has an influence on players' behavior and interaction. Cara Wrigley [20] defines visceral hedonic rhetoric in the context of interactive products as "the properties of a product that persuasively elicit the pursuit of pleasure at an instinctual level of cognition."

II. GAPS

Cara Wrigley [20] describes how the visceral hedonic rhetoric model or tool is relevant to product designers to devising interactive products that will create a continuing emotional impact with the consumers. The whole gamut of design elements of a product elicits some kinds of emotional responses from the users. The researcher mentions that this tool needs validation. It is important to investigate which specific attributes of the visceral rhetoric will give rise to hedonic responses. Furthermore, she recommends that the model could be tested out in other design domain for designers to better understand how consumers' visceral responses are affected by visceral hedonic rhetoric. Research has demonstrated mixed results when it comes to perception of usability and visual aesthetics. Many studies have been conducted to evaluate users' emotional responses in games or to developing tools to measure user experience, however, little is known how the visual rhetoric of a game interface affect user behavior, performance, emotional responses and player experience. There are several gaps identified when it comes to perception of beauty and usability as they relate to user interface design. *"Game interface is part of the surface layer for game control. The experience of the controls is directly tied to the emotional experience of play, thus user interface in a game is very important*" explains video game designer, Robin Hunicke. Another researcher defines aesthetics as the "*sensuous qualities, the emotions, moods, and experiences*" that occurs while interacting with a product [21]. Aesthetics have become an important topic among HCI researchers lately as it forms an integral part of user experience. In the study "what is beautiful is usable," [22] the researchers found out that the relationship between usability and aesthetics were strongly correlated, which was in line with a previous study in the field of marketing and consumer behavior that explained how physical appearance of a product were related to personality traits [23]. On the contrary, when Hassenzahl [24] conducted a study on MP3 skin players, he did not find any relation between beauty and perceived usability.

Usability may be considered more important when the product is used for an important and time critical task, and less so in "*a more relaxed mode of exploration and discovery"* [25]. Hence it becomes imperative to investigate the connection between usability and aesthetic, when computer games are played in the action or goal mode. According to Hassenzahl, classical aesthetic is different from beauty; he considers classical aesthetic to include visual usability that complements usability of interaction. *"Aesthetics is too broad as a term (modern art has an aesthetic value, but most of the time not beautiful) or we use it as a pretentious replacement of what we actually mean, just because aesthetic sounds more learned. In both cases, beauty is the better term,"* declares Hassenzahl in an email conversation. *"Beauty is synonymous to an affect driven evaluative response to the visual gestalt of a product."* [26]. Hassenzahl [24] further explains that beauty and usability is mediated by "goodness" and there is a low correlation between hedonic and pragmatic. The latter finding in turn reiterates Lavie and Tractinsky's study [27] that reveals the correlation between expressive aesthetics and usability is low.

Thuring and Mahlke [28] conducted two studies using portable audio players to find out whether perceptions of instrumental and non-instrumental qualities were related. He simulated two versions of the portable audio players by varying the level of usability and visual aesthetics such that there were four combinations in all. The results revealed that both the variations of usability and aesthetics had an effect independently on emotional responses (valence and arousal) respectively. Mahlke and Lindgaard [29] suggested that the effect of perceived visual aesthetics on emotional aspects of user experience should be investigated further. The researchers explain that there are challenges such as the "inter individual differences of aesthetics judgments" concerning visual aesthetics in human technology interaction that need to be addressed [29]. Everyone has a different level of perception of beauty. When evaluating visual aesthetic aspect of an interface, it is important to take recourse to *Centrality of Visual Product Aesthetics* (CVPA) scale that measures the degree of visual aesthetics an individual accounts for a product [30] to obtain accurate results. Furthermore, Hartmann et al. [26] state that it is imperative to study the behavior of aesthetically sensitive and non-aesthetically sensitive participants in user populations, especially if the aim is to investigate different level of visual complexity in a user interface design. Hence, aesthetic perception also depends a lot on the background of the user.

When a user is exposed to a web interface, usually this gives rise to a halo effect [31, 26] that affects user behavior on subsequent webpages. For instance, if the visual aesthetic is of high quality, even though there is a slight usability issue, the halo effect will override the glitch, and provide a satisfying user experience. One aspect that needs further investigation is the intensity of the halo effect, and how it relates to the degree of usability and aesthetic value. This finding resonates to Hassenzahl and Ullrich's [25] study, which explains in a critical situation, usability is considered the most valuable in contrast to an exploratory/discovery situation, where usability issue is regarded as secondary. Lindgaard and Dudek [16] conducted a study on websites styles to find out that high aesthetic quality and very low perceived usability gave rise to high user satisfaction. It was concluded from this study, "what is beautiful may not be necessarily perceived to be usable." Yet, Tractinsky et al. [22] state "what is beautiful is usable." The above studies are all contradictory to each other. There is little evidence that "explain how and why users judge a particular design to be more or less aesthetic" [26].

Liu et al. [32] investigated the dynamic difficulty adjustment (DDA) in computer games so that the "mechanism in computer games can automatically be suited to gaming experience of the player's characteristics." Liu et al. (2009) further explained that much research is being carried out to evaluate player's performance that is based on game adaptation. It would be more useful to investigate the "affective state" of the players instead. Katsyri et al. [33] made use of electromyography (EMG) to measure emotional responses of 44 participants to aesthetic background images that has a varying level of valence (unpleasant to pleasant) and arousal (calming to exciting). When a pleasant image was presented alone and along with text (new message), positive emotions were evoked. It was found that high arousal images improved superimposed text recall. They suggest that future research should be conducted to investigate the effects of aesthetic background images in other domains.

III. AIMS

One of the major objectives of this research is to carry out a comparative study to investigate how player experience is affected by the visual rhetoric (graphic, text, animation, sound) of a computer game, when participants play two different kinds of game, *Hard Fun and Easy Fun* respectively [18]. It is relevant to examine how the visual aesthetic, including the attributes of interactivity such as "movement speed, movement range, response speed, concurrency" [34] of the dynamic game user interface will affect players' physiological arousal and valence by using physiological measures, in addition to self-report measurement. This research is important because game designers should understand what user interface elements can optimally enthrall game enthusiasts and that developers should devise specific types of games or platform, depending on the objectives (goal mode or action mode), to make it appealing and enjoyable to the right audience. This research will eventually contribute to understand the creation of appropriate user interface components that can lead to better game interaction for players to derive optimal enjoyment and challenges that will lead to enhanced user satisfaction. It is also important to find out how the players' emotional responses are affected by the visual rhetoric of the game interface when they are in different meta-motivational states. Players' emotional responses will be examined in two different conditions: *telic* (goal) and *paratelic* (action). In the goal mode, the participants will be required to play the computer game until they reach a certain score or level, whereas in the action-mode, all participants will be given equal amount of time to play the game. "Usability is considered to be a more important component when the product is utilized for an important and time critical task, and less so in a more relaxed mode of exploration and discovery." [25]. Does this hold true in a game scenario when one plays a *hard fun* and *easy fun* game category respectively? How does the visual rhetoric and gestalt affect the player when he is in a goal mode and in an action mode of gameplay respectively? Moreover, the users can be further segmented into two different groups called verbalizers and visualizers [35] to examine how each group would process and elicit emotional responses towards the visual rhetoric of the game user interface.

Admittedly, components like engagement, immersion, and flow have been used as attributes to measure UX in games. For this study, it is essential to evaluate user experience in games from a multi-dimensional perspective, by appropriating CUE [28] Model and compare the results using other game enjoyment and UX models in the field of Human Computer Interaction Design, like Gameflow [8], Game Experience Questionnaire (GameLab, Eindhoven University of Technology) and UX model such as PIFF [2], that integrates components like Adaptation (Presence and Involvement) and Flow (cognitive evaluation and emotional outcomes), Component of User Experience Model. In addition, it may be useful to compare the above experiments in different cultural settings to check if the results are generalizable.

IV. METHODOLOGY

It is clear that measuring emotional responses in HCI has been a challenge for many researchers despite numerous techniques that have been devised. Physiological (skin conductance, heart rate, pupillary dilation, blood pressure, respiration, vocal tone, facial expressions, muscles action, electroencephalography (EEG), electromyography (EMG)) [36] and self-report tools (pre and post-test questionnaires, interviews, focus groups) have been commonly utilized to measure emotional responses. The advantage with physiological measures is that the researcher can capture emotions accurately. The disadvantage is that it is costly to administer and collect data. Researchers can still detect different kinds of affect using the self-report measurement that includes rating scales, questionnaires, graphical differentials, and verbal depiction of emotional states. The drawback of self-report methods is that the participant is conscious when his emotion is being measured; self-report methods are not suitable for measuring low intensity emotions, according to Frijda [39]. Researchers have commonly utilized self-report measurements because they are cheap to conduct. For the scope of this study, two UX models (CUE and PIFF) will be compared in a game context, and data will be collected when participants play a touch screen game on ipads. Instruments that will be used to collect data include both physiological (skin conductance) and self-report tools.

Data will be collected using the following instruments: (i) Verbalizer-Visualizer questionnaire [37] for screening participants. (ii) HEP Questionnaire [38] for gathering usability data. (iii) Classical and Expressive Aesthetics [27] (iv) Game Experience Questionnaire (GEQ) to measure enjoyment in games. The GEQ is a verbal emotional tool that can measure human experience of media enjoyment, in the context of gaming. The GEQ consists of three modules: actual game experiences, post-game experiences, and social roles of players. The items are based on concepts like immersion, flow, challenge and affect. The stimuli used to collect data are a *hard fun* and *easy fun* category game devised for the touch-screen that can be customized in terms of its usability and aesthetics features. A panel of game experts will first validate the game to ensure that it meets the norms as far as game play, game mechanics and game user interface design are concerned. The game will then be modified into four different conditions as shown in Table 1. The independent variables are usability and aesthetic quality. The dependent variables are emotional reactions and user experience.

TABLE I.

		Usability (A)	
		Low (A1)	***High (A2)***
Aesthetics (B)	***Low (B1)***		
	High (B2)		

Both Qualitative and Quantitative methods will be used in this study. 200 participants will be recruited, and devised into two groups known as “verbalizers” (better at processing text) and “visualizers (better at processing images).” The verbalizer-visualizer [37] questionnaire is a five items questionnaire in which the participant needs to select between the two approaches of processing information, either words or pictures. Based on the visualizer-verbalizer questionnaire, participants will be screened into two groups: the verbalizer and the visualizer, each group will consist of 100 participants. Gender balance will be taken into consideration within each condition. A two-way ANOVA in-between subject test will be performed, with two independent variables: usability (low and high) and visual aesthetics (low and high). In addition, by making use of the CVPA (Centrality of Visual Product Aesthetics) instrument, individual differences among participants can be analyzed in order to understand to which degree users are aesthetically oriented. The data gathered from the CVPA questionnaire can be added as a covariate to the analysis, which will result in an ANCOVA.

The following research questions will be attempted: (i) How is player experience affected in the following states: (telic) goal mode and (paratelic) action mode? (ii) What impact does the level of visual aesthetic in each game category has on the two group of players (visualizers and verbalizers)? (iii) How is the Flow Graph (Challenge v/s Skills) affected by varying levels of the two independent variables? How do components like enjoyment, player experience, player performance and emotional responses affected by the two levels of independent variables?

V. CONCLUSION

This study is expected to answer multiple research questions that will be important for game designers to consider and apply during game development. It is necessary to understand the characteristics of players for instance, whether verbalizers derive more enjoyment than visualizers in gaming context, and how the visual rhetoric affects each category of players. Game designers will need to further explore specific elements of the user interface that give rise to particular emotional responses, in the case of each game category–hard fun and easy fun. It is also necessary to investigate whether the “flow” concepts still hold true when the game attributes (visual aesthetics) are varied in each game condition. Answering the above questions will help pave the way to devise optimal game experience for appropriate audience.

VI. REFERENCES

[1] J. Pine and J. Gilmore, *The experience economy: work is theatre and every business a stage*. Cambridge, MA: Harvard Business School Press, 1999.

[2] J. Takatalo, "Psychologically-Based and Content-Oriented Experience in Entertainment Virtual Environments," Ph.D. dissertation, University of Helsinki, Finland, 2011.

[3] S. Mahlke, "User Experience of Interaction with Technical Systems," Ph.D. dissertation, Technical University of Berlin, Germany, 2008.

[4] "Game Player Data, Entertainment Software Association." Accessed Oct 03, 2012. http://www.theesa.com/facts/salesandgenre.asp

[5] J.Takatalo, J. Hakkinen, J. Kaistinen, and G. Nyman, "User Experience in Digital Games: Differences Between Laboratory and Home." Simulation & Gaming (August 10, 2010). doi:10.1177/1046878110378353.

[6] D. Machaj, C. Andrews, and C. North, "Co-located Many-player Gaming on Large High-resolution Displays." In 2009 International Symposium on Social Computing Applications (SCA-09), 697–704, 2009.

[7] K. Gilleade, A. Dix, and J.Allanson. "Affective Videogames and Modes of Affective Gaming: Assist Me, Challenge Me, Emote Me." In Proceedings DIGRA 2005, 16–20. Vancouver, Canada, 2005.

[8] Penelope Sweetser and Peta Wyeth. "GameFlow: A Model for Evaluating Player Enjoyment in Games." ACM Computers in Entertainment 3, no. 3 (2005).

[10] K. Isbister and N. Schaffer, In Game Usability: Advice from the Experts for Advancing the Player Experience. Burlington, MA: Morgan Kaufman, 2008.

[11] A. Kankainen, "UCPCD: User-centered Product Concept Design." In Proceedings of the 2003 Conference on Designing for User Experiences, 1–13. New York: ACM, 2003.

[12] J.Preece, Y. Rogers, and H. Sharp. Interaction Design: Beyond Human-Computer-Interaction. New York: Wiley, 2002.

[13] A. Kukulska-Hulme, "Mobile Usability and User Experience." *In Mobile Learning: A Handbook for Educators and Trainers*, ed. Kukulska-Hulme & Traxler, (United Kingdom: Routledge, 2005).

[14] R.L. Hazlett, "Measuring Emotional Valence During Interactive Experiences." 1023. ACM Press, 2006. doi:10.1145/1124772.1124925.

[15] P. Zhang and G.M. Von Dran, "Satisfiers and Dissatisfiers: a Two-factor Model for Website Design and Evaluation," *Journal of the American Society for Information Science* 51 (2000): 1253–1268.

[16] G. Lindgaard and C. Dudek, "What Is This Evasive Beast We Call User Satisfaction?" *Interacting with Computers* no. 15 (2003): 429–452.

[17] D.A. Norman. "Emotion and Design: Attractive Things Work Better." Interactions 4, no. 9 (2002): 36–42.

[18] Nicole Lazarro, "Why we play games: Four Keys to more emotion without story," (2004), http://www.xeodesign.com/whyweplaygames.html

[19] A.M. Federoff, "Heuristics and Usability Guidelines for the Creation and Evaluation of Fun in Video Games" (Master's thesis, Indiana University, 2002). http://melissafederoff.com/heuristics_usability_games.html.

[20] C.J Wrigley, "Visceral hedonic rhetoric" (PhD diss., Queensland University of Technology, 2011). http://eprints.qut.edu.au/46040/

[21] I. Engholm, "The Good Enough Revolution – the Role of Aesthetics in User Experiences with Digital Artefacts." Digital Creativity 21, no. 3 (2010): 141–154.

[22] N. Tractinsky, A.S. Katz, D. Ikar, "What is beautiful is usable" Interacting with Computers no. 13 (2000), 127–145.

[23] K. Dion, E. Berscheid, and E. Walster, "What is beautiful is good," *Journal of Personality and Social Psychology* 24(3), 1972): 285-290.

[24] M. Hassenzahl, "The interplay of beauty, goodness and usability in interactive products," *Human-Computer of Interaction*, 19 (2004): 319-349.

[25] M. Hassenzahl and D. Ullrich, "To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals," *Interacting with Computers* 19*,* Issue 2, (2005): 429-437.

[26] Hartmann, Jan, Alistair Sutcliffe, and Antonella De Angeli. "Towards a Theory of User Judgment of Aesthetics and User Interface Quality." ACM Transactions on Computer-Human Interaction 15, no. 4 (November 1, 2008): 1–30. doi:10.1145/1460355.1460357.

[27] T. Lavie and N. Tractinsky, "Assessing dimensions of perceived visual aesthetics of web sites," *International Journal of Human-Computer Studies* 60 (2004): 269–298.

[28] M. Thuring and S. Mahlke, "Usability, Aesthetics and Emotions in Human Technology Interaction," *International Journal of Psychology* 42(4) (2007): 253-264.

[29] S. Mahlke, and G. Lindgaard. "Emotional Experiences and Quality Perceptions of Interactive Products." *In Human Computer Interaction*, ed. J.Jacko, 164–173. 1st ed., (Berlin Heidelberg:Springer-Verlag, 2007.

 ISBN: 978-1-61208-250-9

[30] P.H. Bloch, F.F Brunel, and T.J. Arnold, "Individual differences in the centrality of visual product aesthetics: concept and measurement," *Journal of Consumer* Research 29, no. 4 (2003):551-565. doi: 10.1086/346250

[31] L. Deng and M.S. Poole, "Affect in Web Interfaces: A Study of the Impacts of Web Page Visual Complexity and Order," *MIS Quarterly* 34 no.4 (2010): 711-730.

[32] Liu, Changchun, Pramila Agrawal, Nilanjan Sarkar, and Shuo Chen. "Dynamic Difficulty Adjustment in Computer Games Through Real-Time Anxiety-Based Affective Feedback." *International Journal of Human-Computer Interaction* 25, no. 6 (August 12, 2009): 506–529. doi:10.1080/10447310902963944.

[33] Kätsyri, Jari, Niklas Ravaja, and Mikko Salminen. "Aesthetic Images Modulate Emotional Responses to Reading News Messages on a Small Screen: A Psychophysiological Investigation." *International Journal of Human-Computer Studies* 70, no. 1 (January 2012): 72–87. doi:10.1016/j.ijhcs.2011.08.007.

[34] Y. Lim, S. Lee and K. Lee, "Interactivity Attributes: A New Way of Thinking and Describing Interactivity." 105–108. New York: ACM Press, 2009.

[35] Massa, Laura J., and Richard E. Mayer. "Testing the ATI Hypothesis: Should Multimedia Instruction Accommodate Verbalizer-visualizer Cognitive Style?" *Learning and Individual Differences* 16, no. 4 (January 2006): 321–335. doi:10.1016/j.lindif.2006.10.001.

[36] Shami, N. Sadat, Jeffrey T. Hancock, Christian Peter, Michael Muller, and Regan Mandryk. "Measuring Affect in Hci." 3901. ACM Press, 2008. doi:10.1145/1358628.1358952.

[37] Mayer, Richard E., and Laura J. Massa. "Three Facets of Visual and Verbal Learners: Cognitive Ability, Cognitive Style, and Learning Preference." *Journal of Educational Psychology* 95, no. 4 (2003): 833–841. doi:10.1037/0022-0663.95.4.833.

[38] H. Desurvire, M.Caplan and J.A. Toth. "Using Heuristics to Evaluate the Playability of Games." In CHI '04 Extended Abstracts on Human Factors in Computing Systems. New York: ACM Press, 2004.

[39] Frijda, N.H (1986). The emotions. Cambridge: Cambridge University Press.

Emergent Design System
Using Computer-Human Interactions and Serendipity

Akira Kito, Yuki Mizumachi
Graduate School of Science and Technology
Keio University
Yokohama, Japan
fa0kitou@hotmail.co.jp, yuki-mb@jcom.home.ne.jp

Koichiro Sato, Yoshiyuki Matsuoka
Faculty of Science and Technology
Keio University
Yokohama, Japan
Koichiro_Sato@a3.keio.jp, matsuoka@mech.keio.ac.jp

***Abstract*— This paper describes a basic study on an emergent design system in which serendipity occurs from interaction between computer and human. Serendipity is a natural ability to unexpectedly make interesting or valuable discoveries. The possibility of generating new design ideas will increase if we can utilize serendipity. Therefore, we propose an emergent design system that produces serendipity by using form organizing phenomenon seen in nature and three-dimensional modeling like clay-modeling. Then, we perform elementary experiments with designers. Thus, this system prompts the chances of getting inspirations and unexpected discoveries in the process of deriving ideas. As a result, we show the possibility of generating new design ideas by using this system.**

Keywords-emergence; design system; interaction; serendipity

I. INTRODUCTION

The design process can be roughly divided into two categories: the early process which consists of conceptual and basic designs, and the late process which consists of detailed design [1]. In the early process, novel and diverse design ideas must be obtained from a global solution search under unclear design conditions. Thus, in previous study, we propose an emergent design system. This system is based on the concept of emergence, which is a natural phenomenon that creates diverse organism. We demonstrate diverse design proposals are derived using the system [2]. On the other hand, designer's experience and knowledge earned through trial and error are the key elements in designing. When highly-experienced designers generate design ideas using a representation method, such as a sketching or clay modeling, sometimes unexpected values are discovered by chance. This ability to find something interesting or valuable by coincidence is called serendipity [3]. A previous study on serendipity reported that design ideas not originally envisioned by the designer emerge when designers are devoted to sketching [4]. Therefore, design proposals derived by the emergent design system should enable designer's serendipity to occur.

In this research, an emergent design system which uses computer-human interactions is proposed to support idea generation. This system consists of two fundamental functions: a self-organizing function based on the concept of emergence, and a three-dimensional modeling function imitating clay modeling. During the iterations of these functions, designers should be inspired by self-organizing design proposals. Therefore, generating novel and valuable design ideas will become easier. Additionally, we conduct elementary experiments to test the effectiveness of this system. As a result, we confirm the possibility that the proposed system has ability to support the idea generation.

The paper's structure is presented as follows: first, we describe about the concept of emergence, and propose an emergent design system. Then, we describe about the methods and the results of elementary experiments. Finally, the results are examined and our conclusion is presented.

II. EMERGENCE AND EMERGENCE DESIGN

In nature, various organisms exist in the same environment. In the fields of biology and ecology, scientists have hypothesized that various species have been created through the process of emergence. The concept of emergence is as follows: a new function, the character, and the action acquired by an interactive dynamic process where global order appears by local interactions between individuals, which behave autonomously, with the environment. On the other hand, this order restrains the behavior of an individual [5]. Herein, the appearance of global order is a bottom-up process, whereas the process of restraining individual behavior is a top-down process.

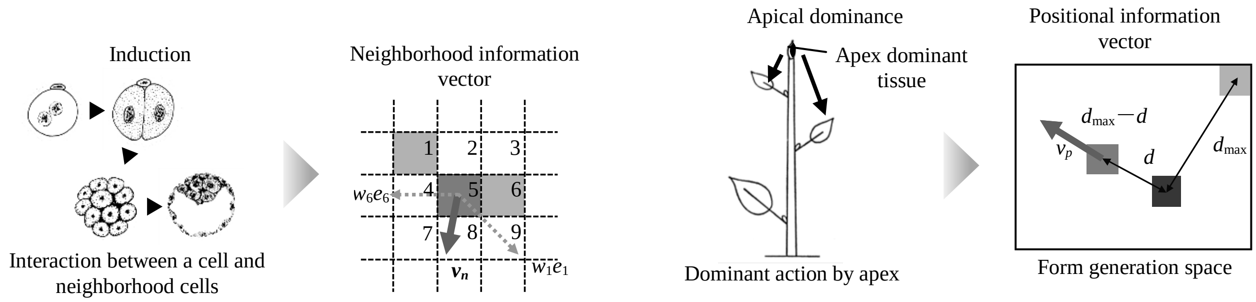
There are two similarities between the early design process and the emergence process. First, the process to generate design ideas through evaluation using certain standards is similar to the bottom-up process which generates the entire feature by the interaction of autonomous components in emergence. Second, the process to optimize detailed parts of the design proposal is similar to the top-down process, which binds the components by entire feature in emergence. Thus, the concept of emergence may be applicable to design, and diverse novel design proposals can be derived by "emergent design" where bottom-up and top-down processes interact.

III. EMERGENT DESIGN SYSTEM

In this section, first the previous emergent design system is described. After that, an emergent design system which uses computer-human interactions is proposed.

A. Previous Study in Emergent Design System

The emergent design system consists of bottom-up process and top-down process. In the bottom-up process,



(a) Induction (b) Apical dominance

Figure 1. Model of the input vectors

diverse design proposals meeting the low standard set by the designer are derived by self-organizing, while the top-down process satisfies the constraint conditions, and optimizes proposals that satisfy the constraints. This system derives diverse designs by going through these two processes.

1) Bottom-up process: In the bottom-up process, forms are generated self-organizationally using Cellular Automata (CA) [6]. In the incidental method, the states of cells in the lattice are updated following a local rule. More specifically, at time t, the state of an element is S_t and the state of the neighborhood (mostly Von Neumann neighborhood or Moore neigborhood) is N_t. The state of the element at time t+1 describes as

$$S_{t+1} = f\,(S_t, N_t) \tag{1}$$

where, f is the transition function which influences the behaviors of the elements.

In the emergent design system, the diversity of an organism is noted, and rules refering to two propoerties for diverse organism morphogenesis, "induction" and "apical dominance", are the input vectors for the CA [7].

An organism is formed by interactions between neighborhood cells. These neighborhood cells affect each other, causing a cell to change and exhibit different features (Figure 1(a)). This property is called 'induction'. The first input is defined as the neighborhood information vector $\boldsymbol{v_n}$, which is expressed as

$$\boldsymbol{v_n} = \sum_{i=1}^{26} b_i w_i \boldsymbol{e_n} \tag{2}$$

where, i is the surrounding element number, b_i indicates the existence or non-existence of an element (1 or 0), w_i is the coefficient of the vector direction, and e_i is the unit vector of the direction to the object element.

In the developmental process, a certain tissue dominates, such as the bud of a plant or the head of an animal. Such tissues are called the apex, and the dominant action by the apex is called 'apical dominance' (Figure 1(b)). The second input is defined as the positional information vector $\boldsymbol{v_p}$, which is expressed as

$$\boldsymbol{v_p} = (\,d_{\max} - d\,)\boldsymbol{e_d} \tag{3}$$

where, d_{max} is the distance between the apex and the most distant cell from the apex, d is the distance between the apex and the object element, and $\boldsymbol{e_d}$ is the unit vector of the direction to the object element. Moreover, the form operating parameter k is set and input vector $\boldsymbol{v_{in}}$ is defined as expressed in Eq. (4).

$$\boldsymbol{v_{in}} = k\boldsymbol{v_n} + (1-k\,)\boldsymbol{v_p} \tag{4}$$

Diverse forms can be generated by changing the form operating parameter k. If the value of k is near unity, then induction tends to strongly influence k. In contrast, if k is near 0, then k is strongly influenced by apical dominance, and a rhomboid or board form tends to be generated. The input parameters in the bottom-up process are the position of apex, form operating parameter k, form generation space, element size, initial element, and evaluation item. The apex position becomes the center of action for apical dominance, and the form generation space is a space that allows CA to be generated. The element size is a voxel and is composed of form. Thus, reducing the element size causes the output to be in a detailed form. The initial element position is where the form generation of CA begins. Thus, diverse design proposals are generated self-organizationally in the bottom-up process.

2) Top-down process: In the top-down process, diverse design proposals generated by the bottom-up process are optimized or modified to satisfy design constraints such as strength or stiffness. For example, the modification method by increasing and decreasing elements, which is inspired by an adaptive function of bone remodeling, is applicable [8].

B. Emergent Design System Using Computer-Human Interactions

We propose an improved emergent design system to support idea generation. Figure 2 shows the flow of the proposed emergent design system. This emergent design system consists of incidental form generation and representation methods. Induction and apical dominance, which are concepts from the previous design system, are

 ISBN: 978-1-61208-250-9

used in self-organizing form generation because these self-organizational concepts enable serendipity.

1) Generate self-organized form in real time: We reconstruct an emergent design system with the Processing programing language, which is benefical to create images, animations, and interactions. In this system, we can see the form generating in real time by increasing elements autonomously. There are two ways to generate forms autonomously. First, selecting a element, and increase elements (Figure 3(a)). Second, elements which increased at last step become the trigger of self-organization. These functions should enable designer's serendipity to occur by revealing diverse and incidentally derived forms.

2) 3D-modeling like clay modeling: In this section, we describe the representation method, which designers can use to generate design ideas. Representation methods are used not only to visualize specific design concepts, but also to convey designers intentions and to adjust design concepts. Our research focuses on a three-dimensional modeling, e.g., clay-modeling, to establish a public image of an idea. In this system, forms can be modified by adding or deleting elements (Figure 3(b)). In order to add elements, you should select an element, subsequently elements at Von Neumann neighborhood or Moore nighberhood will increase. By selecting elements, you can delete the elements.

Three-dimensional models provide information such as depth, which is not intuitive in two-dimensional models. Additionally, through trial and error with representation methods, the likelihood of new ideas from chance or serendipity increases. Hence, three-dimensional representation methods can create new design ideas as well as improve visuospatial cognition.

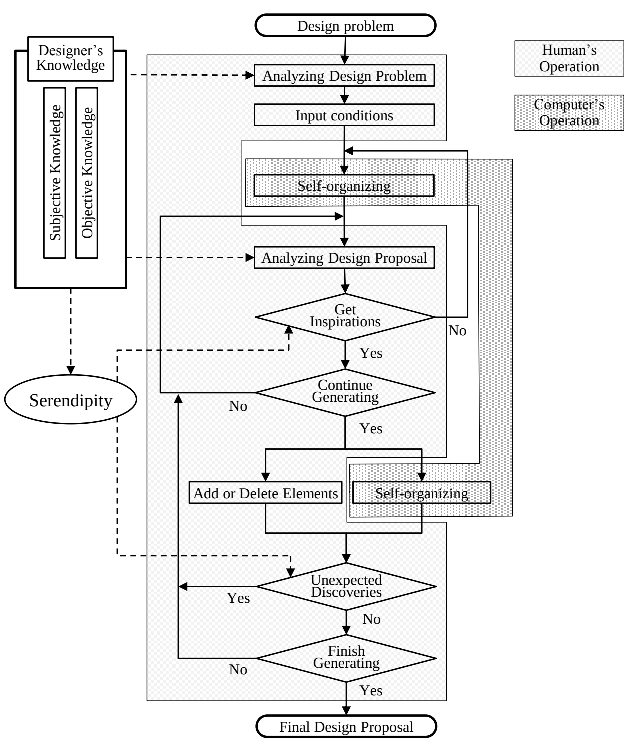


Figure 2. Flow of the emergent design system

IV. Elementary Experiments

In this section, we describe elementary experiments to test the effectiveness of the system.

A. *Experimental Condition and Methodology*

We conduct form generation experiments to clarify the features and effectiveness of the proposed system, including the occurrence of serendipity. In this experiment, we use a chair as a design object. Because a chair design must consider structural features of material mechanics or mechanical dynamics as well as industrial design, a chair is an appropriate application for a basic study to evaluate design idea generation. Four subjects, who are professional designers, participated in the study.

Design proposals are derived using introduced functions via two different patterns. In pattern 1, an initial form, which means the basic form of idea generation, is generated autonomously, and a design proposal is derived using three-dimensional modeling. In pattern 2, an initial form is generated autonomously same as well. Then a design proposal is derived using a three-dimensional modeling and self-organizing form generation. Finally, in both patterns, each subjects draw design ideas within their mind induced by using the system.

Herein the form operating parameter k is set at three levels; 0.1, 0.5, and 0.9. For each ratio, two design proposals are derived by each subject. Then, overall 48 design proposals are derived. To analyze the process of deriving design proposals and characteristic forms, the number of times serendipity occurs is counted: the number of times subjects change his design concept influenced by inspirations and unexpected discoveries. For the same reason, we write

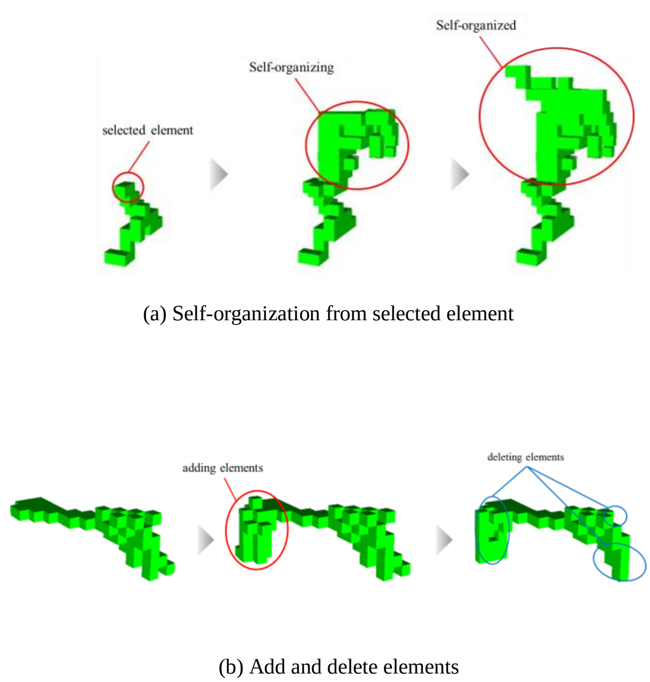


Figure 3. Examples of form generation

down the remarks of subjects while they are generating design proposals. Additionally, we asked the subjects about the usability of this system after the experiments.

B. Evaluation of the Proposed System

Figure 4 shows the examples of the final design proposals derived by the proposed system, and Table 1 shows the number of times serendipity occurs. To clarify the characteristics of the proposed system, we analyze the results from three different perspectives: influence of form operating parameters, influence of form generating pattern, and differences between proposed system and previous system. Analysis from these three perspectives can determine whether the proposed system efficiently supports design idea generation.

1) Difference of form characteristics related to input parameter: To analyze the difference of the form characteristics related to the form operating parameter *k*, design proposals are derived for three different form operating parameters; 0.1, 0.5, and 0.9. Figure 5 shows examples of derived design proposals for each form operating parameter. When parameter is set to 0.1, rod shaped design proposals are derived. In addition to rod shape proposals, aggregated forms are derived when it is set to 0.5. When it's set to 0.9, aggregated forms with numerous elements are derived. Consequently, most design proposals depend on the initial form when form operating parameter is set to 0.9. Similar to the previous system, the value of the form operating parameter controls the form characteristics of the derived design proposals. Thus, the parameter affects the form characteristics of the derived design proposals in proposed system.

When we asked the subjects which form operating parameter provides more inspiration and discoveries, most indicated that it is difficult to judge. However, one subject indicated the process depends on the form operating parameter, confirming the initial form affects the form generation process: increasing elements, or decreasing elements. When form operating parameter is set to 0.9, the system produces aggregated forms, and they are the most geometrically similar forms. Hence, the value of the form operating parameter affects the chance of finding inspiration and discoveries. One method to improve control using the form operating parameter is to restrict the surplus increase in the number of elements. Additionally, one subject suggested changing the form operating parameter while deriving design proposals, indicating that the proposed system needs a new function to change the value of the form operating parameter.

2) Difference of form characteristics related to the form generation pattern: In the experiments, two patterns are used to generate forms. Pattern 1 employs a three-dimensional modeling as a representation method, while pattern 2 uses both three-dimensional modeling and self-organizing.

Figure 6(a) shows the process and final design proposals derived by designer B in pattern 2. Table 2(a) contains the remarks of designer B. Although the proposals are as designer B intended, he does not comment on inspiration or unexpected discovery during the experiment. Similar to Figure 6(a) and Table 2(a), Figure 6(b) and Table 2(b) show the process and proposals and comments of designer B in pattern 2,

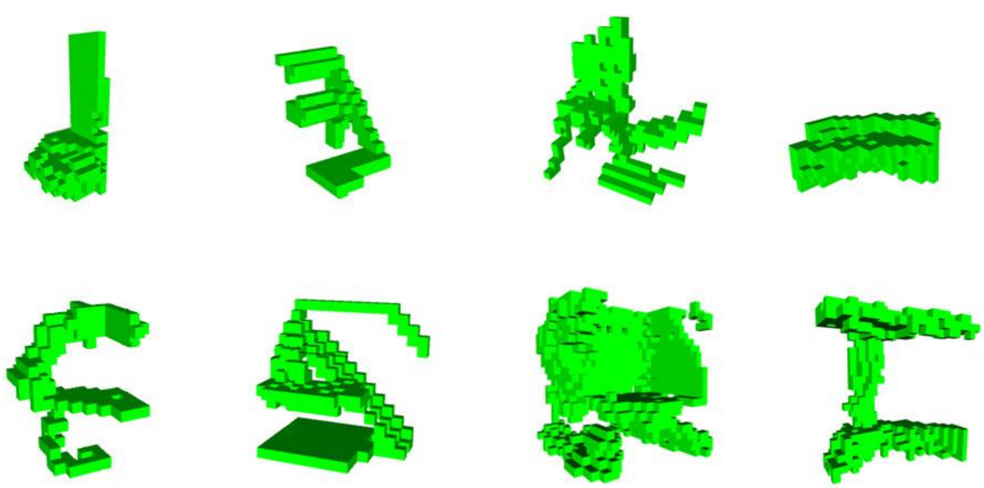


Figure 4. Examples of the derived design proposals

TABLE I. THE NUMBER OF TMES SERENDIPITY OCCURS

		Subject A				Subject B				Subject C				Subject D			
		Pattern 1		Pattern 2		Pattern 1		Pattern 2		Pattern 1		Pattern 2		Pattern 1		Pattern 2	
k	0.1	2	1	1	1	0	0	0	0	2	1	2	2	3	1	5	1
	0.5	0	0	0	0	0	0	1	0	0	0	2	1	0	3	0	0
	0.9	1	1	0	1	0	1	0	0	0	1	0	0	1	1	1	0

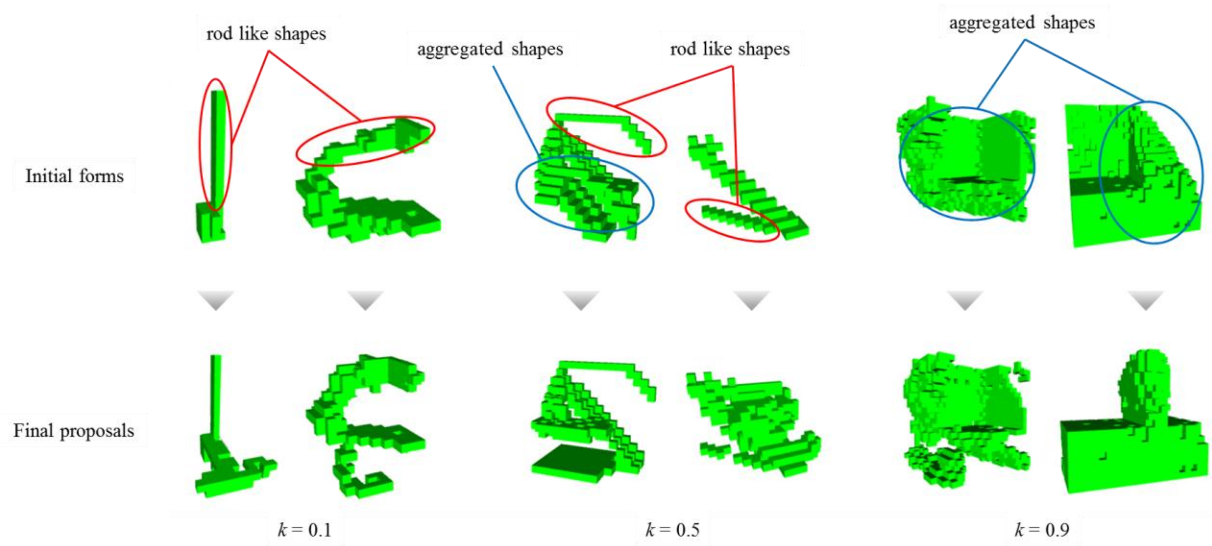


Figure 5. Differences in form operating parameter

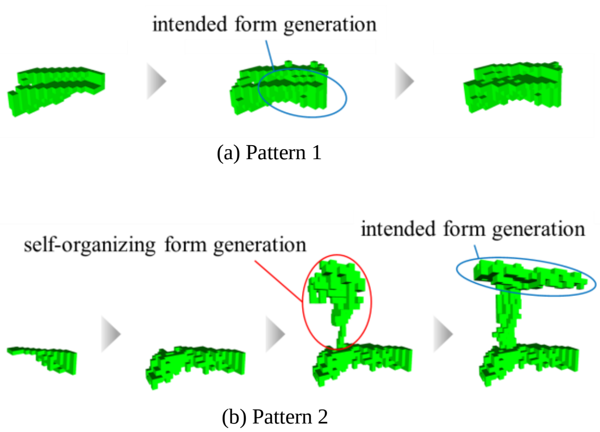


Figure 6. Differences in form generation pattern

TABLE II. REMARKS DURING THE EXPERIMENT

(a) Pattern 1

Process	Remarks
Initial Form	"It's like a bench."
Process	"Need legs for stability."
Final proposal	"Bench at park."

(b) Pattern 2

Process	Remarks
Initial Form	"It's like a bench."
Process 1	"Need more space for several people."
Process 2	"Use this form for the roof."
Final proposal	"Bench with a big roof."

respectively. The inclusion of the self-organized form generation function inspired a new idea, which designer B used to derive a novel design proposal.

We asked which pattern provides more inspiration and discoveries during the experiments. All four subjects answered that pattern 2 provokes inspiration and discoveries. Although there are no significant difference between pattern1 and pattern 2 as Table 1 shows, there is the potential that generating forms self-organizationally yields unexpected novel forms, inspiring new design ideas. Thus, the subjects uses serendipity in pattern 2 because serendipity more likely occurs when using two form generation functions (representation and self-organizing). Hence, results of the experiment and interviews demonstrate that increasing of chance leads to unexpected discoveries.

3) Use of serendipity: Figure 7 shows design proposals of chair derived by the previous system and the sketches of design ideas generated from the proposals. Although designer can generate a design idea by watching the derived proposal, serendipity happens once, if it occurs at all. Figure 8 shows the process of deriving a design proposal by proposed system and the sketch of design idea generated from the proposal. Table 3 shows the remarks during the experiment. The subject used serendipity several times, and various inspiration and discoveries helped generating ideas, confirming that the proposed system allows forms to be generated as designers intended and serendipity emerges during the process.

We asked the subjects which system is easy to generate new design ideas. Half of the subjects answered that the proposed system is easier, but the other half stated that they

Figure 7. Design ideas generated from the previous system

TABLE III. REMARKS DURING THE EXPERIMENT

Process	Remarks
Initial form	"Use self-organization."
Process 1	"Delete elements to make sitting position."
Process 2	"Use self-organization to make a leg."
Final proposal	"An artistic chair."

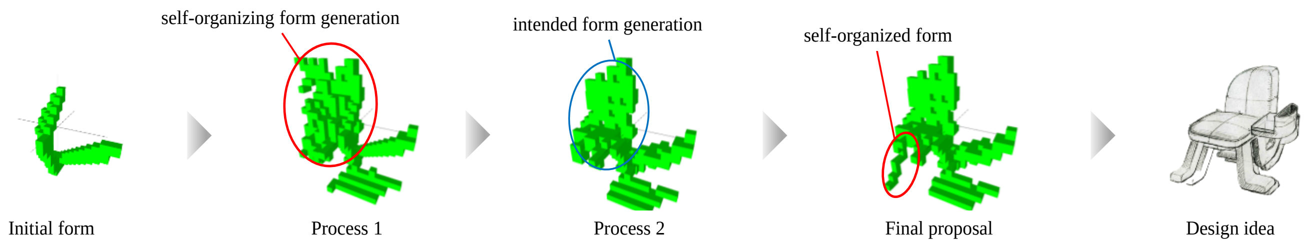


Figure 8. Design idea generated from the proposed system

could not judge. According to designer D, "The proposed system provides chances of finding inspirations and discoveries, generating new design ideas. However, the design idea depends on the final design proposals, which is why it is difficult to say which system is easier to generate new design ideas." Compared to the previous system, the proposed system provides more serendipitous opportunities for inspiration and discovery, but the unfinished design proposals are rejected. Consequently, it is concluded that new design ideas are easier to generate using the proposed system.

V. CONCLUSION AND FUTURE WORK

Herein an emergent design system which uses computer-human interactions is proposed. Additionally, generation experiments to analyze the effectiveness of the system were performed. Compared to the previous system, novel and valuable design ideas are easier to generate in the proposed system. The achievements of this research are described below.

- By enabling interactions in the emergent designs system, designer can generate intended forms via representation method.
- Indicating that both self-organization and representation method help designers to generate novel and valuable design ideas by inspiring designer's serendipity.
- Confirming the possibility that the proposed emergent design system has ability to support the idea generation through iterations of self-organization and representation.

For future researches, we should clarify the mechanisms of how serendipity occurs by studying representation methods which can inspire designer's serendipity. Then, further experiments with larger quantity of subjects to derive statistical consequence should be conducted.

ACKNOWLEDGMENT

This work was supported by the Research Grant of Keio Leading-edge Laboratory of Science & Technology and the Grant in Aid for the Global Center of Excellence Program for "Center for Education and Research of Symbiotic, Safe and Secure System Design "from the Ministry of Education, Culture, Sport, and Technology in Japan.

REFERENCES

[1] K. Sato, Y, Ujiie, Y. Matsuoka, "A Extended Algorithm for Generation of Fibrous Form in the Emergent Design System," ASME 2009 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference(IDETC/CIE 2009), San Diego, CA, USA, DETC2009-86444, pp. 473-480. doi:10.1115/DETC2009-86444.

[2] Y. Matsuoka, "Design science," Maruzen, Tokyo, pp. 82-91, p. 105, 2010.

[3] Y. Maeda, "Hypotheses for Human Potential to Miscellaneous Mysterious Phenomena (2) -Considerations of Synchronicity and Serendipity-," Society for Mind-Body Science, vol.14, no.1, pp.61-65, 2005

[4] T. Sunaga," Why do designers draw sketches?," Special Issue of Japanese Society for the Science of design, vol.12, no.3, pp.12-13, 2005.

[5] S. Kitamura, "Emergent systems-soward a new paradigm for artificial systems. toward system theory of function emergence," Journal of the Society of Instrument and Control Engineers, vol. 35, no. 7, pp. 492–495, July 1995.

[6] A. Ilachinski, Cellular Automata. A Discrete Universe, World Scientific, 2000.

[7] Y. Matsuoka, T. Fujii, "Form-generation method applying properties in the developmental process of organism generating the morphological diversity," Japanese Society for the Science of Design, vol. 49, no. 3, pp. 93-102, September, 2002.

[8] A.Kito, Y. Mizumachi, M. Sekiguchi, K. Sato, Y. Matsuoka, "Modification method by increasing and decreasing elements in the emergent design system," International Conference Advanced Materials Research, vol.422, pp.807-810, 2012, doi:10.4028/www.scientific.net/AMR.422.807.

 ISBN: 978-1-61208-250-9

Sensory Evaluation Method to Create Pictograms Based on Multiplex Sign Languages

Naotsune Hosono
Keio University,
Oki Consulting Solutions
Tokyo, Japan
Hosono903@oki.com

Hiromitsu Inoue
Chiba Prefectural University
of Health Sciences,
Chiba, Japan
cal32840@topaz.plala.or.jp

Yuji Nagashima
kogakuin University
Tokyo, Japan
nagasima@cc.kogakuin.ac.jp

Tomita Yutaka
Keio University
Kanagawa, Japan
tomita@z3.keio.jp

***Abstract*— This paper discusses a method to create pictograms based on multiplex local sign languages with applying the concept of "Context of Use" on dialogue with applying Multivariate Analysis (MVA). Since pictograms are universal communication tools, human centred design (HCD) and context analysis by Persona model are applied. The experiments consist of three steps. The first step is to find out the similarity of a selected word among seven different local sign languages, which are American, British, French Spanish, Japanese, Korean and Chinese by means of MVA. The second step is to create a new common pictogram referring to the first step result by a pictogram designer. The final step is to validate the newly created pictogram by MVA. Under the cycle of HCD, the pictogram designer will perform to summarize the expression of several local sign languages by this method. The acquisition of this experience is to include it as a pictogram design guideline for context of universal communications such as emergency and traveling situations. Through the proposed method, the relationship between selected words and local sign languages are initially explained by sensory evaluation of the subjects. Currently the outcome of pictograms or icons of this experiment are implemented on the modern tablet computers with a touch panel display considering computer-human interactions.**

Keywords- Context of Use; Human Centred Design; Pictogram; Universal Communication; Sensory Evaluation.

I. INTRODUCTION

This paper discusses a method to create pictograms or icons referring to multiplex local sign languages with the concept of context of use on dialogue and Multivariate Analysis (MVA) [1]. Since pictograms or icons are universal communication tools, Human Centred Design (HCD) [2] and context analysis by Persona model by Alan Cooper [3] are applied in this research. This research was started in order to investigate the context of universal communication through local sign languages.

HCD is based on the context of use which is organized by four factors as user, product, task and environment in use (Figure 1). The research scope covers not only linguistic studies of sign language but also HCD with context of use [4].

The structure of this paper is that at first section the research purpose with deaf people issues in the case of emergency is introduced. Then simple and easy to use pictograms or icons are to be required as an efficient communication tool. In the middle sections to find out such pictograms or icons applying sensory evaluation method is discussed. The validity of the newly created pictograms or icons is discussed in the final section.

II. RESEARCH PURPOSE AND ISSUES

The purpose of this research is to figure out a method to create meaningful pictograms or icons referring to several local sign languages [5]. The sign language (SL) is basically a communication method from one person to the other for hearing impaired persons. The main factors of sign language consist of the hand shape, location and movement. There is a dilemma that SL is a language with motion whereas pictograms or icons are still ones. There was quite a discussion among researchers. Then hand shapes and locations are drawn by an animation and movements are done by arrows referring to a snapshot of the related local sign languages.

III. RESEARCH PROCEDURES

Considering such research purpose and issues, the following three steps with seven phase research procedures are prepared;

Step1:
- Phase 1: Determine a concept
- Phase 2: Create Persona Model and its Scenario
- Phase 3: Key words extraction especially on emergency and travelling situations
- Phase 4: Conduct first Sensory Evaluation with seven local sign languages to extract similarities.

Step 2:
- Phase 5: Design a summarized pictogram referring to the result above.

Step 3:
- Phase 6: Conduct second Sensory Evaluation with seven local sign languages adding a summarized pictogram to validate
- Phase 7: Conclude a method

A. *Phase 1: The Context Determination*

Based on the concept described above, two context situations of emergency or travelling have initially been chosen [6]. Alan Cooper proposed the Persona Model [3] related to HCD where several situation representing Personas are imaginably created in order to simulate and find how they will behave under a certain context. This method is highly accepted by the manufacturers in creating new

 ISBN: 978-1-61208-250-9

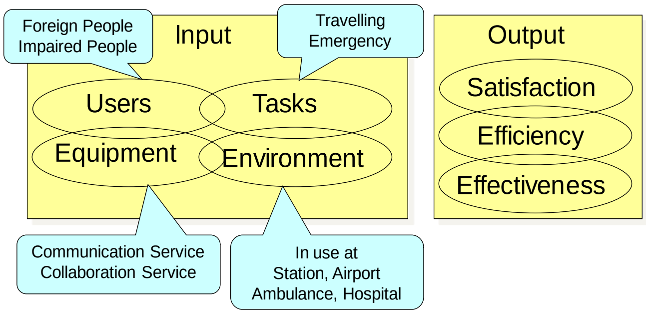


Figure 1. Context of use of Guidance on usability

product plans and has been applied to service science [7].

B. Phase 2: Persona Model and Scenario Creation

The first step is to create two Personas with applying the Persona Model under HCD [2]. The first Persona is a deaf person in a situation where he suffers a sudden illness while commuting in the morning, and is carried to the hospital by an ambulance (Figure 2). The second one is an office lady who lives in Hong Kong and has to visit Tokyo on business and then pleasure.

Diary like scenarios underlying Personas are described from discussions with colleagues utilizing the Brain Storming Method. These scenarios mainly pay attention to the dialogues between the Persona and those people surrounding [8]. The first scenario of the deaf person in an emergency consists of about 600 words (equivalent to 3000 Japanese characters) and second with the traveling woman about 1700 words (equivalent to 8500 characters).

C. Phase 3: Key words extraction on situations

This research is focused upon dialogues with several participants and referring to observations from the view point of the provider and the receiver under dialogue principle [8].

The next phase is to extract words that are fundamentally essential to the dialogues of the scenarios. 37 words were selected and categorized by discussions with colleagues.

Looking at the dialogues in the scenarios under the selected context, the hardest process is initiating the dialogue to a stranger. In modern times, people are worried about security. They are extremely cautious when approached by an unfamiliar person. Several interjections are included to assist the initiation of dialogues.

D. Phase 4: First Sensory Evaluation with seven local sign languages

The research is initially focused upon creating pictograms or icons to make dialogues since the fundamentals of sign language are hand shape, location and movement. This research references to a collection of

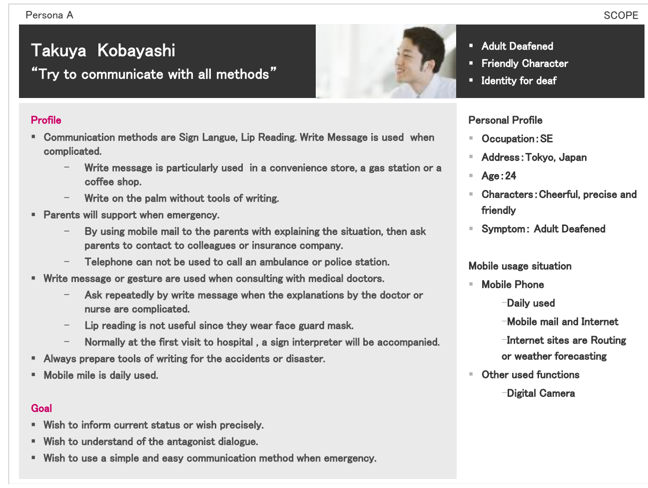
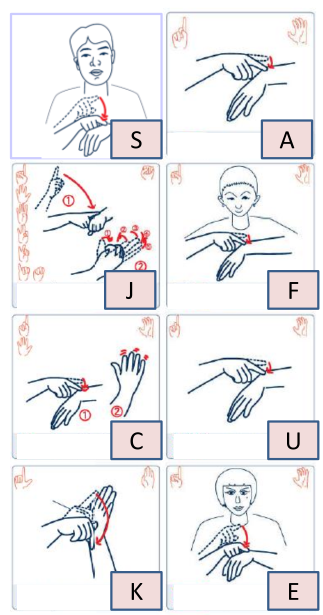


Figure 2. An example of Persona model

animation figures consists of seven local sign languages whose author is a deaf architect, gave overwhelming support to the research by supplying and permitting reference to the database. The seven local sign languages are of American, British, Chinese, French, Korean, Japanese, and Spanish [9].

In the experiment, subjects are first shown an expression with the collection of animation figures consists of seven local sign languages. After then subjects are informed of the sign meaning, they are requested to vote with 19 tokens which of the seven different local sign language expressions (samples) best coincides with the informed image. They are asked to put all 19 tokens on the condition that they are permitted eventually zero voting on some samples.



Abbreviations:

A: American Sign Language (ASL)
J: Japanese Sign Language (JSL)
F: French Sign Language (FSL)
C: Chinese Sign Language (CSL)
B: British Sign Language (BSL)
K: Korean Sign Language (KSL)
E: Spanish Sign Language (SSL)

Figure 3. A sample of voting sheet of

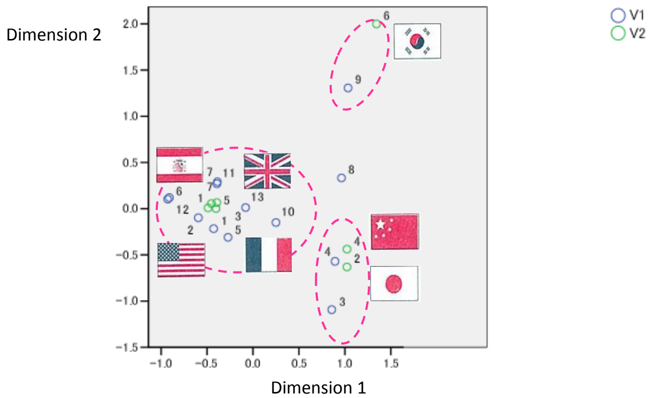


Figure 4. A plot of "When ?" with seven sign languages.

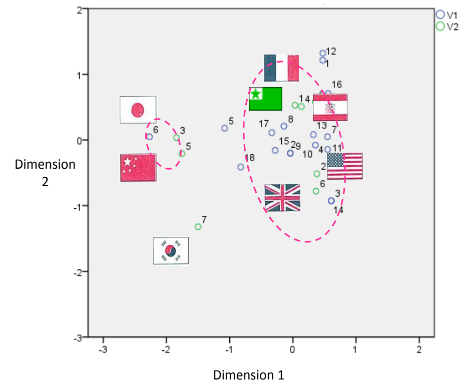


Figure 5. A supplementary treatment plot chart of "When ?" with eight sign languages

This sensory evaluation method can easily make relative comparisons between the seven expressions of local sign languages and is more applicable than the ordering method or pair comparison method. An example of voting sheet of "When ?" with local sign languages is shown in Figure 3 except the pictogram "S".

Then the correspondence analysis of Multivariate Analysis (MVA) by statistic software; Statistical Package for Social Science (SPSS) [10, 11] is applied. The outcome is plotted as similar local sign languages are to be plotted closely on a plane. In the characteristics of correspondence analysis, the subjects who have general and standard ideas are positioned in the centre, whereas those who have extreme or specialized ideas are positioned away from the centre. The center crossing point of the first and second Eigenvalues is gravity point or average point.

The first experiment subjects are 13 people in their 20's including nine science course students, four humanity course students. Some have experience living overseas and sign language interpreting. After voting by the tokens, all the subjects are asked of their confidence level with Semantic Differential (SD) method [12].

Figure 4 is an example of outcome chart where "When ?" is plotted.

E. Phase 5: Summarized Pictograms Design

Analyzing a plot chart of "When ?" in Figure 4 with seven sign languages, many subjects feel that French, American, British and Spanish sign languages are representing similarly "When ?" since they are plotted closely on the chart.

Following to the cycle process of HCD, the original designer is asked to summarize and design an animation like ictogram showing exclusively French, American, British and Spanish by referring to the outcome by the sensory evaluation mentioned above. Then the newly designed pictogram, which is "S" in Figure 3, is added to seven local sign languages with American, British, Chinese, French, Korean, Japanese, and Spanish.

F. Phase 6: Second Sensory Evaluation with seven local sign languages with summarized pictogram

The next procedure is the same manner as the first experiment of Phase 4. After subjects are informed of the sign meaning, this time they are requested to vote with 23 tokens which of the eight different local sign language expressions including newly designed one which is a pictogram "S" in Figure 3 will be the best coincides with their image. The procedure was the same manner as the first sensory evaluation step of phase 4, and the correspondence analysis of Multivariate Analysis (MVA) by SPSS is once again performed. The outcome including the newly designed pictogram is plotted with other seven local sign languages in order to prove and measure whether the newly created pictogram represents of the cluster.

The second experiment subjects are 20 engineering department students in their 20's including two female students. Almost all except three are different subjects from the first experience. After voting by the tokens, all the subjects are again asked of their confidence level with Semantic Differential (SD) method.

Figure 5 is an example of outcome chart where "When ?" is plotted. Comparing Figure 4 with Figure 5, Figure 5 is plotted under uncertainty for factor rotation. The newly designed one will represent French, American, British and Spanish sign languages since it is plotted close to those sign languages. Whereas Japanese, Korean, and Chinese plotted further down.

In order to prove the outcome, Supplementary Treatment of MVA by SPSS is applied with adding newly designed one to the seven sign languages. These deployments of the plots are similar in seven and eight sign languages experiments.

G. Phase 7: Conclude the Method

Through the experience of the first Sensory Evaluation with seven local sign languages of 37 words, many sign language expressions are identified by representing the meaning. Among them the most converged seven words of "when?", "good-by", "When ?", "thank you", "where?", "toilet", and "expensive" among 37 are selected by means of brain storming.

Comparing two outcomes of Phase 4 (Figure 4) with seven local sign languages and of Phase 6 (Figure 5) with eight local ones, followings are concluded.

- Selected seven newly designed animation pictograms are all positioned in the centre of the related local sign languages cluster.
- Even though almost of the subjects are different at the first and second experiment, the general outcome plot patterns hold similar patterns in space.
- In western sign languages of French, American, British and Spanish tend to be plotted closely together.

IV. CONCLUSION AND FUTURE WORK

This paper discusses a method to extract the summarized expression of several local sign languages in order to draw pictograms or icons by applying the sensory evaluation with MVA. The experiments consist of three steps.

The first step is to find out a pictogram is a majority common expression upon a word among seven local sigh languages. Looking at the first step, this method looks valid in practice since French and American sign languages are similar by historical background, and in fact and they are plotted close to each other. The second step is to prove the characteristics of the pictogram represent the meaning of the word. The final step is to validate the newly created pictogram by MVA. Almost all of the newly designed pictograms positioned in the centre of the cluster then it is representative of the clusters.

In the example of "When ?" among seven sign languages, analyzing by the Supplementary Treatment of MVA, the newly designed pictogram will be representing French, American, British and Spanish sign languages since it is plotted close to those sign languages on the plane.

Currently the outcome of pictograms or icons of this experiment are implemented on the modern tablet computers with a touch panel display considering computer-human interactions.

V. DISCUSSIONS

Through the proposed method, the relationship between selected words and local sign languages are initially explained by sensory evaluation by the subjects. Under the cycle of HCD, the pictogram designer will perform to summarize the expression of several local sign languages by this method. The acquisition of user experience is to include it as a design guideline for instance of the context of emergency and traveling situations.

The issues are that the quality of the newly designed pictogram depends on the designer's ability to summarize several ones. The newly designed pictograms are still biased by sign languages in this research in order to become much easier communication tool, and require further improvement to simplify and easily to understand for everybody.

Considering the results of the second experienced phase to prove the outcome design by Supplementary Treatment of SPSS, the proposed method is one of the guidelines to create pictograms by referring to several sign languages.

ACKNOWLEDGEMENT

A collection of local sign language database is supplied and permitted for research use by Mr. M. Akatsuka of Architectural Association of Japanese DEAF (AAJD) as well as Mr. M. Suzuki who is a deaf architect and proposed idea.

REFERENCES

[1] Hosono, N., Inoue, H., Tomita Y.: Sensory analysis method applied to develop initial machine specification, Measurement, vol. 32, pp. 7-13 (2002).

[2] International Organization for Standardization: ISO9241-210 (former 13407:1999), Ergonomics Human-centred design processes for interactive systems (2010).

[3] Cooper A.: About Face 3, Wiley (2007).

[4] Miki H., Hosono, N.: Universal Design with Information Technology (Japanese version), Maruzen (2005).

[5] Horton W.: The Icon Book, John Wiley & Sons, Inc. (1994).

[6] International Organization for Standardization: ISO9241-11, Ergonomic requirements for office work with visual display terminals (VDTs), Guidance on usability (1998).

[7] Hosono N., Miki H., Suzuki M., Tomita Y., Urgent Collaboration Service for Inclusive Use, HCII2009, in CD-ROM (2009).

[8] International Organization for Standardization: ISO9241-110, Ergonomic requirements for office work with visual display terminals (VDTs), Dialogue principles (2006).

[9] Akatsuka, M.: Seven sign languages for tourists: Useful words and expressions, Chinese-Japanese-American Working Group (2005).

[10] Field, A.: Discovering Statistics Using SPSS 3rd edition, Sage Publications (2009).

[11] SPSS: Categories in Statistical Package for Social Science ver.18, SPSS (2009).

[12] Hosono N., Inoue H., Miki H., Suzuki M., Nagashima Y., Tomita Y., Yamamoto S., Service Service Method to create Pictograms referring to Sign Languages, HCII2011, in CD-ROM (2011).

Accessibility and Augmented Reality into Heritage Site Interpretation

A pilot experience with visitors at the monument Lonja de la Seda in Valencia (Spain)

Marina Puyuelo Cazorla, Lola Merino Sanjuán, Mónica Val Fiel
Escuela Técnica Superior de Ingeniería del Diseño
Universitat Politècnica de València UPV
Valencia, Spain
e-mail: mapuca@ega.upv.es, mamesa@ega.upv.es, movalfie@ega.upv.es

José Luís Higón Calvet
Escuela Técnica Superior de Arquitectura de Valencia
Universitat Politècnica de València UPV
Valencia, Spain
e-mail: jhigon@ega.upv.es

***Abstract*— This paper summarizes a pilot experiment with Augmented Reality (AR) at this architectural monument declared World Heritage Site by UNESCO in 1996. The project aims to increase accessibility to this site and provide users, in this case, visitors to the place, with intuitive experience of this technology. The experience and the application are proposed as a complement to the visit in the real environment, emphasizing the site as a context for situated learning that contributes to knowledge. Comprehensive visits to world heritage sites endow them with extraordinary cultural and social value. Augmented reality was chosen to activate the visit experience of the place by providing elements that can be visualized and manipulated directly by users in the real environment of the building itself without the need for any excessively invasive equipment in this historical and artistic context. This application is proposed firstly, to resolve perceptual issues due to poor lighting, the distance from many details and access to some areas and secondly, to explain some of the more complex construction aspects. This pilot experiment aims to establish initial contact with visitors to the place in order to obtain relevant information for modifications and adjustments to improve the components designed and developed for this site.**

Keywords; augmented-reality; accessibility; inclusive-design; heritage-sites; interactivity.

I. INTRODUCTION

Information and communication technologies together with technological tools for graphic and tactile representation are currently undergoing continuous development and expansion in the form of a wide variety of emerging applications in objects with multiple added characteristics. These new applications and their future possibilities open up a wide panorama for research ranging from the perceptual experiences these tools can offer to the nature of future devices that will incorporate them to interact with people.

This research presents an Augmented Reality application, based on a graphic substrate, to improve accessibility to this built heritage site. The most significant challenge in this experiment is make this technology available to visitors to the monument so they can explore and visualize building details as directly and intuitively as possible. This pilot experiment attempts to test on visitors, potential users firstly, receptivity to the whole experience and secondly, the use of the equipment which forms the system's general interface.

Augmented Reality is defined [1] as the combination of the real and the virtual, interactive in real time and can also be said to integrate the real context. With Augmented Reality it is possible to overcome certain limitations in the environment and/or users, and provide new perceptions and understanding of these places. The choice of AR technology for this site is based on a previous accessibility study [2] and the idea that this virtual reality application is less intrusive and provides the greatest opportunity for users to directly manipulate complex three-dimensional models in real space environments [3]. Augmented Reality (AR) is characterized as a technology that enables users themselves to interact with and visualize virtual models.

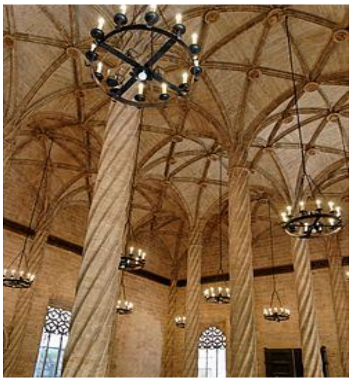


Figure 1. This pilot experiment took place in the room of columns at the Silk Exchange, Valencia. This room is the most charismatic space in the building and it characterised by its slender twisted columns.

This paper shows the role of this particular technology as mediator of a virtual representation/reality in building new ways of interacting with places and knowledge that can be useful in situations of disability.

This tool enables a new relationship to be established between the interpretation and conservation of heritage property, permitting access to simulated perceptions that helps to overcome the common limits of visual accessibility in a specific place, due to distance, high ceilings, poor lighting or in visitors themselves: impossibility of moving towards or approaching details, reduced discrimination and visual identification abilities, etc.

The first part of this paper presents some direct antecedents of this type of application and experiments and their possibilities for accessibility. At the heart of the communication presents the AR system design and main components created for this pilot experience at this singular building. We observe interaction with users in the real scenario, their choice of the element to be observed, handling the marks and their appraisals. Finally, conclusions can be drawn firstly, from users' experience of the tool in order to correct aspects of the components which can contribute to more intuitive use. Secondly, from the experience as a contribution to accessibility to the place and its contents. To this end, visitors who had used AR to visualize some elements and understand them better completed an in situ questionnaire after visiting the site. The questionnaires were not analyzed but will be taken into account to guide the redesign of the data gathering tool itself. The redesigned questionnaire will be used in a longer implementation experiment at this site when this project has finished.

II. Virtual reality applications for interpreting and visiting monuments

The basis for this approach is situated learning [4] which assumes that learning occurs more efficiently in a real context and that the real context becomes an important part of the basic knowledge associated with this type of learning. The use of virtual reality systems can provide interpretations, stimuli and explanatory models that adapt more closely than others to the discourse of the disciplines to which they refer. In terms of accessibility, both in situ and through virtual experiences, they have an undisputed ability to compensate, by providing perceptions for understanding and enjoyment of such places, and are obviously irreplaceable in certain situations of disability.

A. Augmented Reality

AR is a field of unceasing experimentation as a technological tool at the service of three-dimensional graphic representation under constant development and expansion through a wide range of very different applications. Its ability to expand information in real environments through easy-to-use dynamic virtual visualizations can solve, among others, aspects of visual accessibility in real time [5].

Many AR applications generate different types of environments with varying requirements and spatial performance. Some of these tangible or tactile AR applications use a backlit table which can be manipulated from the surface by manually dragging elements as though it were a desktop [6]. In a particularly interesting approach, the MIT's Tangible Media group implemented interfaces linked to tangible objects in an attempt to demonstrate that interaction is more intuitive and functional when a relationship is established between the physical objects being manipulated and augmented and the computer-generated image (Ishii & Ullmer, 1997). SitePack is directly related to the area of representation and visualization of architectural models and warrants mention because of the instrumental nature of the application [7]. It was designed for exteriors and enables the combination of dynamic virtual 3-D elements to assess their impact on the real environment. Similarly, the Prisma 22 project proposes visualization of real landscape space by superimposing additional virtual content with relevant information. The portable AR21 application attempts to explain the historical evolution of heritage buildings using animations and virtual descriptions in situ.

III. A pilot experiment with Augmented Reality at the Lonja Silk Exchange in Valencia

This section describes the first experimental application of AR designed for this site to explore the use and coexistence of the system in this singular monument as context scenario. The research focuses on interaction with users in the real environment where visitors can experience the application, identify the elements and acquire broader knowledge of the place and certain details.

This pilot experiment aims to establish initial contact with visitors to the place in order to obtain relevant information for modifications and adjustments to improve the components designed and developed for this site. These are: the marks created for visitors to identify and use, the number of elements in the experiment, each of the 3-D models created and introduced in the AR engine and which are visualized in movement on the screen. Finally, a questionnaire was used to obtain users' opinions on receptiveness, use and appraisal of the application as an accessibility tool. In this initial contact with users and visitors to the place, the aim is to obtain relevant information for any modifications and adaptations required to improve the components designed for the site.

The method for creating this Augmented Reality experience has been developed in two stages. First, the technical design of the 3-D visual models, based on selected elements from the site, and the panels which are the basic interface for expanding the chosen motifs. Second, the application was implemented in the real scenario for visitors to the building la Lonja.

A. Technical desin: the 3-D models and the markers

3-D visual models were created based on selected elements from the site. There are two different types of chosen elements according to volume: fully three-dimensional construction elements (Fig 2), such as the shaft of the column, the capitals and stairs and plan-dependent elements such as windows traceries (Fig 3), doors, coffering and paving.

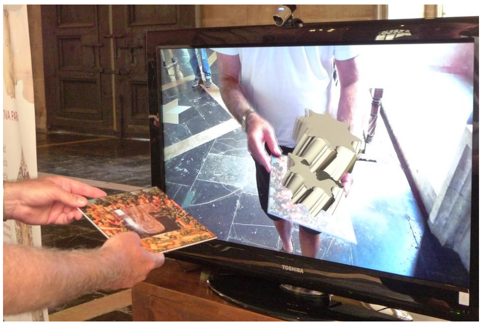


Figure 2. The model of the column is another of the elements that provide a better explanatory model of the construction.

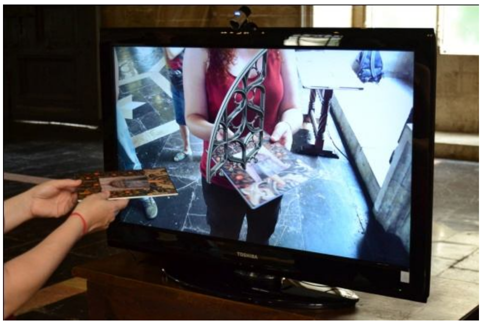


Figure 3. The proportions of the gothic traceries are particularly light and delicately related to the visualization format.

After determining the elements to be shown, three-dimensional models with representation capacity must be generated to be visualized in the AR application. These models basically consist in geometry and texture or surface [8] and there are two possible approaches [9] to generating them as three-dimensional models: data acquisition using laser scanner technology or, three-dimensional modeling of the elements based on their graphic definition.

The large quantity of descriptive graphic information on this building, together with general and partial measurements for the elements to be modeled, was considered sufficient. The models thus obtained have a low triangular load because the geometry representing them clearly and unmistakably defines the edges and apexes of the models.

To represent the textures of the models, shadows were pre-calculated and then mapped on the model as material. Shadow pre-calculation used as surface material is not intended to be a faithful imitation of the texture of the represented objects, but to conceptualize the different materials in the constructive definition. The results for model geometry and texture are sufficiently representative to show the relevant aspects of the models on the AR application.

Once the virtual 3D models where installed in the RA engine, the next decisive step for usability is to design the element that will act as intermediary in the user's experience: the markers. This interaction panels are used to trigger the loading of the different 3D models.

Considering the informative superiority of iconologies over the written word to identify actions in other interactive systems, familiar figurative references can be versatile resources for an interaction space.

The choice of graphic metaphors can help users to understand interface functionality better and improve accessibility. In an attempt to follow these considerations and bearing in mind the characteristics of this historic building, we didn’t use common binary marks and it was proposed that these marks should be identified directly by users and related to the motifs they show. So, graphic formats were designed ex profeso for this site, paying particular attention to integration and comprehension. This process required some images to be retouched and the proportion of representation areas to be varied to obtain those that were finally used in the experiment.

The outcome, assessed by users, is a composition focused on a square combining photographic images with high iconic and representative value. The image that identifies the element to be visualized is in the centre, framed by another photographic image that refers to the period and what was worked with in the building, that is, silks.

These panels provide the necessary texture to be read by webcam. The proportions, size and format of these marks were pre-tested to ensure robust webcam recognition, and levels 4 and 5 were found to be optimum for the tracking technology employed. This system uses LabHuman markerless library which is based on the works of Wagner et al. [10] and Kim et al. [11]. The proposed method works fine with textured planar objects, that is, with planar images which contain a great number of keypoints. This library implemented a method based on image retrieval techniques in order to recognize simultaneously a great number of images, specifically, Nister and Stewenius [12] vocabulary tree method was used.

For this experience, 6 different panels were executed and tried. 4 of these panels have two different scales of visualization which allows more detailed information. To get this zoom in or zoom out effect user must move the markers closer or farther to the webcam.

Figure 4. Marks designed and used to activate 3D models of different details in the building.

B. *Usage Scenario and Pilot Study*

A total of 45 individuals from different age groups and background who were visiting the monument with or without a guide, took part in this pilot experiment and completed all parts of it. The site has a constant stream of visitors throughout the year, with peak times during the summer and the cruise season (cruise ships have been stopping at the port of Valencia since 2008). At those times especially, some three thousand visitors can concentrate around the building on a single day, with a predominance of families and people over fifty. Because the project aims to validate AR

 ISBN: 978-1-61208-250-9

technology as an accessibility resource, we were particularly interested in inviting participation from adults and the elderly.

The used scenario in which the experience was developed consisted of an information panel with a description of the experience to be done and the technical equipment for sampling: support table, 23 inch television screen with web cam, laptop loaded with the models, the six marks printed in matte finish surface with the elements to be visualised representing the models. It is worth noting that, given the characteristics of this protected heritage site, one of the requirements was include the minimum number of objects in order not to interfere with the global view of the monument.

The fact that users see their own hands on the screen means they immediately experience and interpret what has to be done while they direct the panel to the camera. The speed at which users learn to handle the marks and remember and find positions and the shapes they want obeys the concept of “proprioception” [13] and can be observed in these immersion experiments.

Another positive aspect worth mentioning [14] is that despite having visited the building or just passing through, visitors were willing to try out different models and dedicate more or less time to each one as they chose. They never said they were tired of standing up or of having to get up and direct the panels towards the video camera. Although the effect of surprise disminished over time trying different models, it persisted and increased at multiple occasions depending on the 3Dmodel.

Figure 5. Some visitors like the elderly showed a lot of interest and surprise when using the application, observing elements that are difficult to see because of where they are (a). The opportunity to locate different elements in the place and identify them in the real space is one of the greatest attractions for users and a game for young children (b).

After the visit to the monument, users were asked to fill a survey in an informal pilot study in order to gather qualitative data. These interviews were conducted individually when they had used the marks and the AR application as they wished. Participants were asked about the degree of satisfaction with the use of this technology and how AR helped them to perceive aspects of the monument, which otherwise they had gone unnoticed. A general overview of the results suggests that users are satisfied with the use of technology, finding easy and useful for enhancing the comprehension/perception of the monument.

Finally it is worth noting that they showed less interest in this conventional participation instrument than in their prior experience with the application. In general, we found that almost all participants felt comfortable using the system, it made an impact on them and they valued the use of AR in real contexts like this one very positively.

IV. Conclusion and Future Work

This experience with Augmented Reality at this heritage site shows the effectiveness of the tool as an instrument for accessibility and communication of the specific cultural content of the place by actively supplementing and extending vision. Visitors' experience with this technology has provided a profile of component characteristics for adaptation to the target public from the perspective of inclusive design.

The marks were found to be effective and appreciation, impact and interest was found to be linked to their aesthetic and figurative dimension. Users enjoy identifying and selecting the motifs they are going to visualise with this technological tool. They can select what they wish to "enlarge" from the whole visit. This aesthetic and figurative quality increases cognitive compatibility in relation to these signals and consequently, is a determining factor for improving interaction with the system.

The use of this technology acts in certain publics with a “picklock” effect for certain segments [12] who are directly attracted by these communication systems such as children and young people who feel enveloped, comfortable and participative with this type of digital media. Children and young people were eager and skilled in using the marks and tried to explore each model in great detail as if it were a game. However, other visitors closer to the target population of the experience as accessibility resource took part and were very attracted by the idea of being able to visualise the details of the place better. For these users, the fact that the system was easy to use and did not require the use of common computer peripherals was, as they noted, an advantage of the system

The nature of these visual applications closely related to holistic perceptual phenomena makes it difficult to obtain quantitative results [8]. As perceptual phenomena are subject to interpretation, analysis and observation is complex. There has been no attempt to extract quantitative results from the experiment through the satisfaction questionnaire because, although it is a suitable strategy for obtaining information on aspects of technology use and acceptance, it has only been analysed in order to rectify the data collection instrument itself, sound out types of users, their receptiveness, satisfaction, etc.

As a pilot experiment, it has enabled us to profile particular aspects in order to improve the application and its implementation in the Lonja Silk Exchange, such as redefining certain characteristics of the models, their visualization on screen and their evocation from the corresponding mark.

The user's position in front of the screen had to be determined and highlighted and the camera re-located to obtain a medium shot of users during their experience to make it more global and inclusive. Groups of 3D models were established according to proximity in order to favour their localization in the place and details in the satisfaction questionnaire were reconsidered to make it shorter and more operational in this use context.

We conclude that this AR application project will enhance this artistic and historical heritage site in Valencia, giving an interactive resource which increases visitor interest, its participation and a holistic experience of this site.

The future works of the research team are directed to improve the aspects of usability of the tool, allowing thus a better access to the heritage to the groups with limitations in his visual or perceptive capacities.

ACKNOWLEDGMENT

This pilot experience has been carried out in collaboration with the HumanLab as part of a research project in progress, funded by the Universitat Politècnica de València Research Program PAID0511 "Nuevas Aplicaciones de las Tecnologías Gráficas para la Mejora de la Sostenibilidad, el Conocimiento y la Accesibilidad al Patrimonio" (Ref 2786).

It is also necessary to be grateful to the Town hall of Valencia which has allowed us to realize the experience in this protected building.

REFERENCES

[1] R. Azuma, "A Survey of Augmented Reality", Presence: Teleoperators and Virtual Environments, Vol. 6, nº 4, 1997, pp. 355-385.

[2] M.Puyuelo, L. Merino, M. Val, F. Felip and J. Gual, "Access to World Heritage Sites: Design Products that Transform Sites into Collective Spaces for Enjoyment and Interactive Learning", DESIGN PRINCIPLES AND PRACTICES: AN INTERNATIONAL JOURNAL in Champaign, Illinois, USA by Common Ground Publishing LLC http://ijg.cgpublisher.com/product/pub.154/prod.364 (último acceso 23/01/2013), 2010, pp. 409-433.

[3] D. Schmalstieg and A.Fuhrmann, VRVis Research Center for Virtual "The Studierstube Augmented Reality Project", *PRESENCE:* Vol. 11, No. 1, February © 2002 by the Massachusetts Institute of Technology, pp. 33–54.

[4] S. Scribner, "Thinking in action: Some characteristics of practical thought", en Sternberg, J. y Wagner, R. K., Practical intelligence: Nature and origins of competence in the everyday World, Cambridge University Press, Cambridge, 1986, pp. 13-30.

[5] G. E. Jaramillo, J. E. Quiroz, C. A. Cartagena, C. A. Vivares, and J. Willian Branch: Mobile Augmented Reality Applications in Daily Environments". Revista EIA, ISSN 1794-1237, Número 14, Escuela de Ingeniería de Antioquia, Medellín (Colombia), Diciembre 2010, pp. 125-134.

[6] B .Ullmer and H. Ishii, "The metaDESK: Models and prototypes for tangible user interfaces". Proceedings of ACM User Interface Software and Technology (UIST'97), 1997, 223–232.

[7] M. B. Nielsen, G. Kramp, and K- Grønbæk, "Mobile augmented reality support for architects based on feature tracking techniques". Proceedings of the Workshop on Interactive Visualization and Interaction Technologies, Krakow, Poland, 2004.

[8] O. Bergig, N. Hagbi, J. El-Sana, K. Kedem and M. Billinghurst, "In-Place Augmented Reality", Virtual Reality Springer-Verlag London Limited, 2011, pp. 201–212.

[9] F. Fantini, "Variable Level of Detail in Archaelogical 3D Models Obtained through a Digital Survey". Revista EGA Nº19, 2012, pp 306-317

[10] D. Wagner, G. Reitmayr, A. Mulloni, and T. Drummond, Schmalstieg D, "Real-time detection and tracking for augmented reality onmobile phones", IEEE Trans Vis Comput Graph 16(3), 2010, pp. 355–368.

[11] K. Kim, V. Lepetit, and W. Woo, "Scalable real-time planar targets tracking for digilog books", Vis Comput 26(6), 2010, pp.1145–1154

[12] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree", In: IEEE Comput. Soc. conf. comput. vis. pattern recognit., vol 2. IEEE, 2006 pp 2161–2168.

[13] M. Mine, F. Brooks Jr., and C. Sequin, "Moving objects in space: Exploiting proprioception in virtual-environment interaction". Proceedings SIGGRAPH '97, 1997 19 –26.

[14] C. Baber, H. Bristow, S. Cheng, A. Hedley, Y. Kuriyama, M-Lien, J. Pollard, and P. Sorrell, "Augmenting museums and art galleries", 13rd Conference on Human-Computer Interaction Interact 2001. Tokyo, Japan (9-13 July, 2001, pp. 439-447.

[15] O. Bergig, N. Hagbi, J. El-Sana, K. Kedem, and M. Billinghurst, In-Place Augmented Reality, Virtual Reality Springer-Verlag London Limited, 2011, pp. 201–212.

A Deported View Concept for Touch Interaction

Alexandre Alapetite,
Henning Boje Andersen
Department of Management Engineering,
Technical University of Denmark
Produktionstorvet 424,
DK-2800 Kongens Lyngby, Denmark
{alal, hebq}@dtu.dk

Rune Fogh
Centre for Playware,
Department of Electrical Engineering,
Technical University of Denmark
Elektrovej 325,
DK-2800 Kongens Lyngby, Denmark
{rufo}@dtu.dk

Ali Gürcan Özkil
Department of Mechanical Engineering,
Technical University of Denmark
Produktionstorvet 426,
DK-2800 Kongens Lyngby, Denmark
{alio}@dtu.dk

***Abstract*—Following the paradigm shift where physical controls are replaced by touch-enabled surfaces, we report on an experimental evaluation of a user interface concept that allows touchscreen-based panels to be manipulated partially blindly (aircrafts, cars). The proposed multi-touch interaction strategy – involving visual front-view feedback to the user from a copy of the peripheral panel being manipulated – compares favourably against trackballs or head-down interactions.**

Keywords-HCI; Tactile interaction; Touch; Blind; Visual attention; Cockpit; In-vehicle systems

I. INTRODUCTION

There is a trend to replace physical controls (buttons, dials, switches, etc.) by touch-enabled surfaces with adaptive layout. Doing so provides some advantages but also some drawbacks such as the loss of convenient blind activation, where users "feel" controls with their fingers. To explore this problem, we have made a study of a "deported view" mode of interaction with touchscreens and we report the results of an experimental evaluation of the concept. This mode of touchscreen interaction can be particularly useful in aviation and automotive applications in which pilots and drivers may be required to maintain a head-up position for safety reasons.

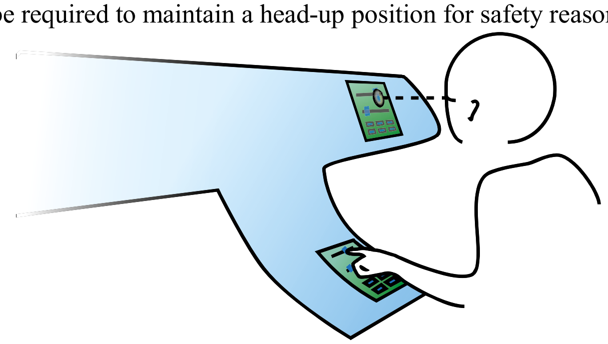


Figure 1. The "deported view" concept in a cockpit.

Figure 1 illustrates the proposed concept: When a user must operate (blindly) a control panel that lies outside his or her natural line of sight, the display area (being touched) will be temporally duplicated on a control screen area in front of the user. Simultaneously, a pointer is shown on the front display indicating the position of the user's finger. By doing so, the lower touch-display panel becomes an indirect pointing device somewhat similar to that of a track-pad.

II. STATE OF THE ART

The idea of implementing touch-based interaction in cockpits dates back to the "Super Cockpit" (US Air Force 1986) which contained a touch-enabled display system to provide *spatial awareness to the pilot in all directions and in three dimensions*. Furness investigated the system and the challenges it poses to the human factors community [1], focusing on *how should information be rendered in the "focal" versus "ambient" visual areas of the display*? NASA also investigated touchscreen input concepts for interaction with a large screen cockpit display as early as 1990 [2], comparing three input methods: thumball, thumb switch and touchscreen, to interact with a large, multi-window, "whole-flight-deck" display. While the thumball concept outperformed the others in simulator tests, it is acknowledged that *touch entry would be useful in transport environment, dependent on error-free operation*. Pilots emphasized the importance of positioning the touchscreen, suggesting that it should be placed so close to the throttle controls that the operator need not reach forward.

Touchscreens were then put into operations, first in military aircrafts such as the Dassault Rafale in 1991 [3]. Since 2009, Garmin [4] and others offer similar touch-based panels for commercial aviation. Kaminani [5] identified two unique interaction issues with touchscreen in cockpits, namely *touch activation and accidental touches,* and *fatigue due to extending arms to touch and reach.*

Similar work has also been done for automotive applications. Young et al. [6] research experiments on using touch interfaces for music selection in cars show how the participants' driving performance decreases and the amount of time drivers have their eyes off the road increases. A multi-touch interface implemented in a car's steering wheel was presented by Döring et al. [7], showing that gestural interactions can significantly reduce visual demand for non-critical interactions. Bach et al. [8] compared three different means of interaction (tactile, touch, and gestures) for in-vehicle systems in terms of efficiency and visual demand, finding that gestures reduce visual demands while touch interaction is most efficient. Although gesture interactions demand less visual attention [8], they are less useful for more complex environments with dozens of precise controls such as found in a cockpit.

 ISBN: 978-1-61208-250-9

The differences in performance between touchscreens and physical interfaces, such as a joystick or a stencil, have also been studied by Kadytė and Tétard [9], showing how applications such as navigating through menus were more efficient with touch devices than the joystick.

Using a touch display for remote control has been investigated for the television market. Han et al. [10] developed a remote control using absolute position on a touch display: the user's finger is detected by infrared when hovering over the remote, which is shown on the TV with a shadow of the finger (pointer). But the hovering concept was found more difficult and stressful than current remotes.

Turning to research into cognitive demands and input modalities, other studies have demonstrated the importance of spatial memory [11] and kinaesthetic cues for user interfaces. For instance, Tan et al. [12] have shown a 19% increase in spatial memory for information controlled with a touch-screen compared to a mouse interface. Similarly, Jetter et al. [13] found that multi-touch instead of mouse input improves users' spatial memory and navigation performance for user interfaces involving panning.

We would like to explore a variation of the concepts above in order to preserve some of the assets traditionally offered by physical interfaces, such as spatial memory, while migrating to touch-screen interfaces.

III. FUTURE AIRCRAFT COCKPITS

The main use-case for the concept presented in this paper is for future aircraft cockpits, as envisioned by the EU project ODICIS [14] on "One Display for a Cockpit Interactive Solution" (2009-2012, see Figure 2).

Figure 2. The ODICIS single-screen tactile cockpit.

It is the view of the ODICIS consortium [14], [15] that the next evolutionary step in cockpit design is to provide the crew with a large continuous, adaptive, multi-touch display that no longer is limited by the physical boundaries of adjacent displays. A single screen offers a more flexible design of the human-machine interface and therefore improved opportunities for providing the right information at the right place at the right phase of flight. Moreover, it offers optimised usage of the main instrument panel in terms of display space. Finally, prompted by novel design concepts in the consumer industry, new tactile technologies are introduced into the cockpit, offering more natural ways of interacting with controls, e.g., by allowing the locus of interaction to coincide with the locus of feedback [16].

The ODICIS concept aims to integrate the emerging tactile technologies into its single display concept. While offering new possibilities, it also raises many questions and potential pitfalls and, therefore, challenges in creating a design to make the most out of this new tool. One of the most salient drawbacks of the approach is the loss of tactile feedback of physical buttons and knobs [16]; hence, pilots will have to rely on other means of ensuring that they operate the intended control and in the intended way. This is especially true for those control-panels, which are located at the periphery of the pilot's centre of vision.

IV. THE "DEPORTED VIEW" CONCEPT

To compensate for the loss of tactile feedback, which can increase the amount of time spent on looking at the interface [6], the "deported view" concept makes use of the flexibility of having a large display (as ODICIS) and instantly duplicates the lower-positioned (or overhead) panel being operated, onto the front display and thus directly in front of the pilot. The feature is optional since pilots can choose to use it or not at any time.

In an aircraft cockpit, the surfaces available in the periphery of a pilot's vision (sides, above) provide important locations for controls and lamps, because many functions require direct access, and not everything can fit in front of the pilot.

The user's initial touch will activate the duplicate screen (the "deported view" panel) and show the area now initialized by the user's finger – like a mouse cursor. The "deported view" control panel will thus function both as a display and pointing device.

The proposed pointing device uses absolute positioning, differing from the relative positioning of a track-pad or mouse. This will allow a user to operate the panel looking either directly on the lower display or using the duplicate on the front display.

Using absolute positioning also means that users can use their spatial skills in order to increase efficiency. While the first finger controls the location of the cursor, clicking is performed with a tap of a secondary finger (Figure 3). In order to avoid inadvertent actions, the lower positioned control panels should not trigger any action when the first finger is just landing. This multi-touch approach creates a mouse-like interaction that should be efficient and robust to operate.

V. EXPERIMENTAL SETTINGS

We report an experimental assessment of a prototype that implements the "deported view" concept, compared with three other modes of interaction: a traditional touch interaction (in two different positions) and a trackball, all programmed with the Microsoft technology WPF (Windows Presentation Foundation). Ten participants were recruited (2 females, 8 males, all but one right-handed; mean age 24 years, range 20-27). The participants executed two tasks simultaneously (Figure 3).

Figure 3. The "deported view" experimental setup, with the operator using a second finger to perform a click.

The first task required participants to move sliders on a screen. The second task was a control task, which required them to respond as fast as possible and within three seconds to a randomly appearing visual cue, and which was introduced to prompt participants to maintain a head-up position as much as possible. Participants would receive points depending on the speed with which they completed each of the tasks, thereby using a game concept as motivation for being as fast as possible. A penalty was implemented for the control task to prevent users from clicking if no cue was presented, discouraging them from gambling when they were looking head-down at the control panel at their side.

Prior to the experiments, participants were trained for each setup until they reached a certain skill level (criterion level). The order of tasks and task training was randomly distributed among participants in order to balance out any effect of order of training, task familiarity, or fatigue. Each setup trial was performed for two minutes.

A. Detailed Description of the Tasks

(Cf. appendix for an explanatory video.)

1) Control Task (left hand)

The *control task* required the participants to respond to a visual alert, in the form of a red box on the left side of the upper monitor (see Figure 3). The alert appeared randomly at intervals ranging between 0.1 to 3 seconds, and was turned off by pressing any of the joystick's buttons.

2) The Slider Task (right hand)

The program would randomly display one of two sliders on either the desktop screen, lower screen or both, depending on the setup. Participants had to pull the slider down as fast as possible to reach a highlighted area and then release it. When completed, the slider disappeared and a new one would be randomly generated. When the participants had completed three sliders, the program would wait three seconds before generating three more sliders.

B. Description of the Four Setups

1) The "Deported View" Setup

As described in the concept description, the lower multi-touch screen was used as remote input for operating the sliders. Users could look down but were instructed that it was more efficient to stay heads-up.

2) The "Direct Touch Desktop" Setup

In this setup the sliders were operated by the touchscreen on the desktop in a traditional touch manner. However, before the sliders were available on the desktop screen, the correct area must be selected on the lower screen by a single tap. This was done to simulate that a secondary control-panel was selected in a lower position, but operated in front of the participant.

3) The "Direct Touch Low" Setup

Participants operate the two sliders directly on the lower touchscreen and must thus look down to be sure to touch the correct areas.

4) The "Trackball" Setup

Similar to the "direct touch desktop" setup, the proper panel must first be clicked on the lower screen, (again to simulate that a secondary control-panel is selected) after which the slider must be operated by trackball. The trackball is located near the lower touchscreen (Figure 4), in a setup similar to what is found in some plane cockpits.

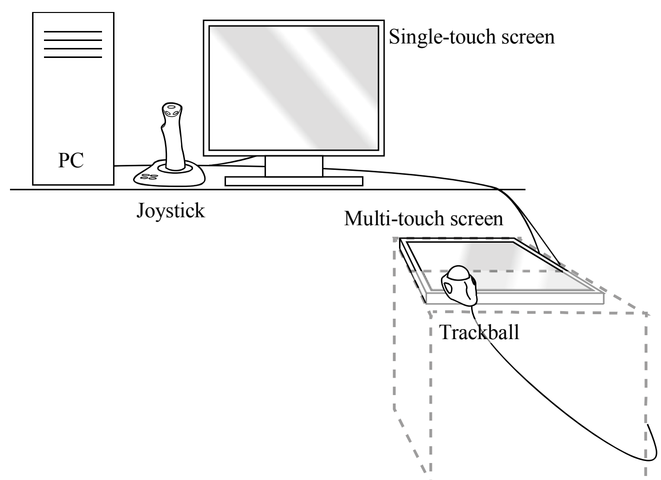


Figure 4. Illustration of experimental setup.

C. Data Collection

The participants' performances were logged in a text file by the program during the experiments. The main performance indicator was the amount of time before the participants responded to the alert ("reaction time"), and how fast they completed the slider tasks ("completion time").

VI. Quantitative Results

The software R version 2.15 was used for the statistical analyses. First, an analysis of variance (ANOVA, within-subjects) showed that there was a strongly significant effect of the type of setup with regard to the "completion time" ($p<0.0001$), and "reaction time" ($p<0.0001$).

TABLE I. POST-HOC BONFERRONI-ADJUSTED PAIRWISE COMPARISON WITH P-VALUES FOR SIGNIFICANT DIFFERENCES.

α<0.05	**Deported view**	**Direct touch desktop**	**Direct touch low**
Direct touch desktop	Completion Time (p<0.004)	-	-
Direct touch low	Completion Time (p<0.0004) Reaction Time (p<0.0005)	Reaction Time (p<0.004)	-
Track ball	Completion Time (p<0.002)	Completion Time (p<0.0001)	Completion Time (p<0.0001)

A post-hoc Bonferroni-adjusted pairwise comparison was then conducted on the two dependent variables with a significant effect ($\alpha \leq 0.05$), as reported in TABLE I.

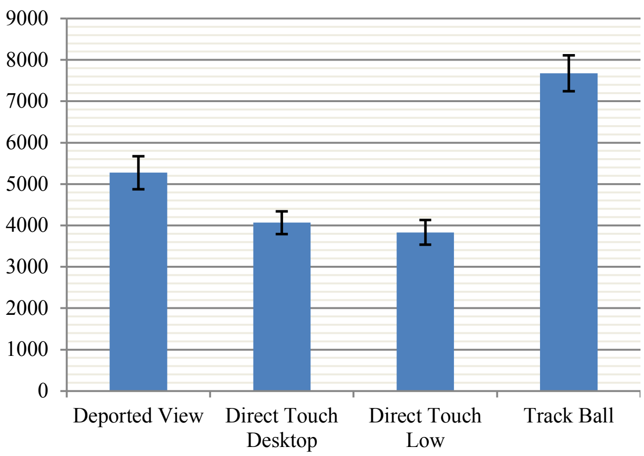


Figure 5. Average time (ms) for participants to complete the task of pulling the 3 sliders for the four different setups. Error bars indicate the standard deviation.

Figure 5 summarizes the results of the participants' average time of completion for pulling down three coherent sliders across the four setups. The average amounts of completed tasks (i.e., three sliders in a row) were: 14.5 for "deported view", 17 for "direct touch desktop", 17.7 for "direct touch low" and 10.9 for the trackball.

The "trackball" is by far the least efficient mode of operation (p<0.004, cf. TABLE I). The "deported view" concept ranks third, being considerably faster than the "trackball" (p<0.0004), while not quite as fast as the two direct touch setups (p<0.004).

Figure 6 summarizes the results of the participants' average reaction time to turning off the alert (the measure of head-up attention). The "deported view" and the "direct touch desktop" setups perform the best, and there is a significant difference (p<0.0005) between "deported view" and "direct touch low". However, there is no significant difference between "deported view" and the two other setups ("direct touch desktop" and "trackball").

Regarding the number of errors during the control task (pressing buttons at a wrong time, or no reaction within 3 seconds), no significant difference between the four setups was found.

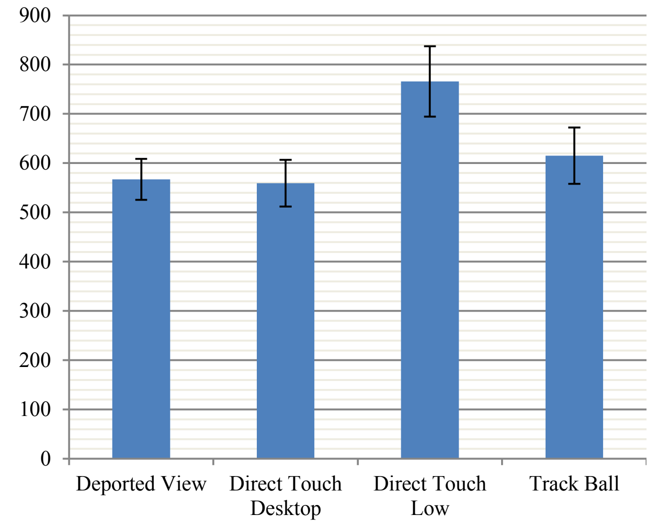


Figure 6. Average time (ms) for participants to react to the control task for the four different setups. Error bars indicate the standard deviation.

VII. RESULTS FROM QUESTIONNAIRES

The participants ranked the four setups according to their general preference and also according to the perceived stress experienced. The results are summarised in Figure 7, where the different setups received 0 points for being ranked last, 1 point for second last and so on up to 3 points.

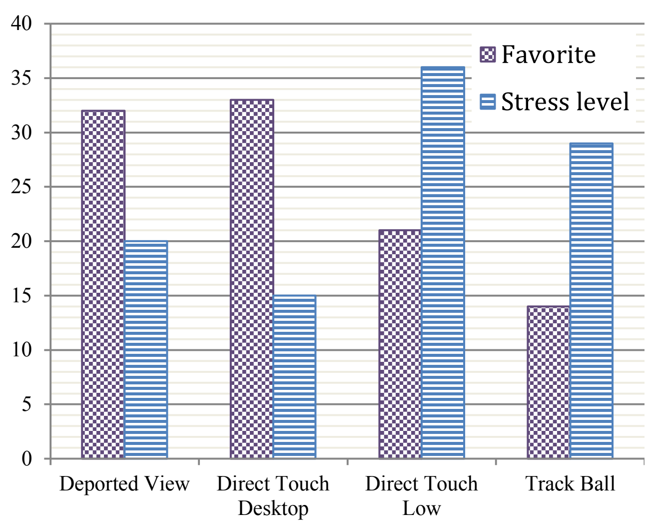


Figure 7. Ranking results from the questionnaire.

The "direct touch low" has the highest reported stress level, while the "deported view" and the "direct touch desktop" were globally preferred by the participants.

VIII. INTERPRETATION

The lower performances of the trackball (Figure 5, TABLE I, p<0.002) may be explained by the fact that a three-step manoeuvre is needed to start the task: first grabbing the trackball (which is smaller than the tactile areas), then move the ball to locate the cursor (not obligatory since the cursor is always visible and placed in the middle of the screen, but a natural behaviour), and finally operate the ball to move the cursor to the appropriate area. This manoeuvre is relatively more complex than with direct touch interaction, which allows these three steps to be performed at once. Although the participants were not familiar with trackballs, they had a chance to train before the experiments just like for the "deported view" and the other setups.

The lower operating speed of the "deported view" compared to the two "direct touch" conditions (p<0.004) can be explained by the fact that the "deported view" uses an artificial feedback (cursor), which both introduces some technical latency, and is more demanding for the user. The physical movements involved in the "deported view", while somewhat similar to a mouse, are not as natural as the "direct touch" ones. Furthermore it was clear that the participants were very accustomed to a relative positioned pointing device (such as a mouse or track-pad) compared to the "deported view" absolute positioning. Using this inappropriate mental model, the participants sometimes tried to move the cursor by moving the finger a bit forward, lifting the finger back to the original position and moving a bit forward again (see Figure 8), like with mouse when reaching the end of the mouse pad. With absolute positioning, this approach will not move the cursor onward but instead keep bringing it back to the start position. Further studies on the effect of using absolute position for an indirect pointing device should be investigated.

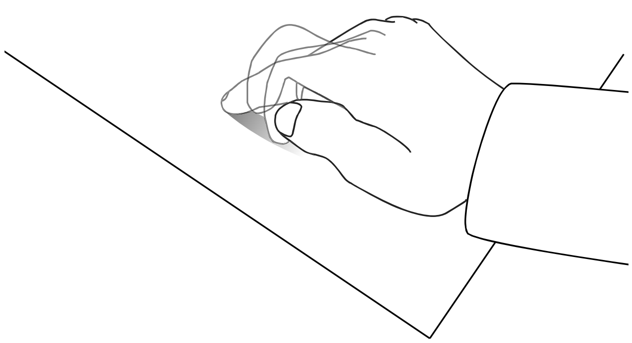


Figure 8. Typical beginner mistake of trying to move the cursor by repeatedly pushing, and lifting the finger back to the original location.

Regarding the reaction time (Figure 6), the fact that the "deported view" performed better than the "direct-touch low" (p<0.0005) setup is because participants could stay heads-up and therefore immediately notice the visual alert appearing on the upper screen.

IX. CONCLUSION AND FUTURE WORK

The results suggest that the "deported view" concept could provide an efficient and secure means for interacting with out-of-sight displays in e.g., automotive and aviation applications.

While the "deported view" design is slower to operate than a touchscreen directly on a front panel, it offers better ergonomics than the latter, as it does not require the user to have an arm in the air with ensuing fatigue [5]. In any case, it is not always possible to place all instruments directly in front of the user at all times due to space and ergonomic considerations. In this regard the proposed concept – temporally showing a secondary touch display in front of the user – may be helpful and safer. Finally, the proposed method proved to be intuitive, attracted greater preference rating and was more efficient than the traditional trackball (which is used in current aircraft cockpits), and resulted in better reaction times than head-down interactions. Anyhow, the concept is mainly meant to supplement and enrich future touchscreen-based solutions, not to replace or compete with other technologies such as haptic feedback, and in combination these solutions may very well benefit from each other to produce increased usability.

This interaction concept might also, we may speculate, be convenient in the tablet and gaming industry. With an increase in integration between TV screens and tablets, the "deported view" concept could provide a means for interacting with the TV screen on the tablet without looking down. In gaming, this could provide some of the many controls needed to play real-time strategy games – which normal gamepads cannot handle – or accessible menus when playing fast paced shooter games. These hypotheses are left for future studies.

While the results are promising, further investigation is required to assess human factors aspects of operational use in a cockpit environment, as well as the effect of different types of feedback such as haptic and auditory.

X. ACKNOWLEDGMENTS

The work presented in this paper has been carried out as part of the EU project ODICIS [14] on "One Display for a Cockpit Interactive Solution" (2009-2012), led by Thales Avionics (France) and partly funded by the European Commission's Framework Programme 7 under grant number ACP8-GA-2009-233605.

XI. APPENDIX

A video of the different setups presented in this article is available at:
http://alexandre.alapetite.fr/research/odicis/#DeportedView

XII. REFERENCES

[1] T. A. Furness, "The Super Cockpit and its Human Factors Challenges," Human Factors and Ergonomics Society Annual Meeting, pp. 48 -52, 1986. doi:10.1177/154193128603000112

[2] D. R. Jones, and R. V. Parrish, "Simulator comparison of thumball, thumb switch and touch screen input concepts for interaction with large screen cockpit display format," NASA Technical Memorandum; no 1025, 1990.

[3] P. Coni, L. Laluque, J. N. Perbet and J. Blanc, “Capacitive Touchscreen Under Electromagnetic Environment,” EuroDisplay’2011 conference, 2011.

[4] S. Pope, “Garmin G3000 brings touch screen tech to flight deck,” Avionics Magazine, 2009.

[5] S. Kaminani, “Human computer interaction issues with touch screen interfaces in the flight deck,” In Digital Avionics Systems Conference (DASC), 2011 IEEE/AIAA, 6B4–1, 2011. doi: 10.1109/DASC.2011.6096098

[6] K. L. Young, E. Mitsopoulos-Rubens, C. M. Rudin-Brown and M. G. Lenné, “The effects of using a portable music player on simulated driving performance and task-sharing strategies,” Applied Ergonomics, 43(4) 738-746, 2012. doi:10.1016/j.apergo.2011.11.007

[7] T. Döring, V. Gruhn, D. Kern, P. Marshall, M. Pfeiffer, A. Schmidt and J. Schöning, “Gestural Interaction on the Steering Wheel – Reducing the Visual Demand,” Proc. of the 2011 annual conference on Human factors in computing systems (CHI'2011), ACM, pp. 3033-3042, 2011. doi:10.1145/1978942.1979010

[8] K. M. Bach, M. G. Jæger, M. B. Skov and N. G. Thomassen, “You Can Touch, but You Can’t Look: Interacting with In-Vehicle Systems,” CHI’2008, pp. 1039-1048, 2008. doi: 10.1145/1357054.1357233

[9] V. Kadytė and F. Tétard, “The role of usability evaluation and usability testing techniques in the development of a mobile system,” NordiCHI’2004, 2004.

[10] J. Han, G. Lee N. Lee, S. Choi and W. Lee, “Remote Touch - Touch-Screen-like Interaction in the TV viewing Environment,” CHI’2011, pp. 292-402, 2011. doi:10.1145/1978942.1978999

[11] A. Cockburn and B. McKenzie, “Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments,” Proceedings of CHI'2002, the SIGCHI conference on Human factors in computing systems, 2002 doi:10.1145/503376.503413.

[12] D. S. Tan, J. K. Stefanucci, D. R. Proffitt and R. Pausch, “Kinesthetic cues aid spatial memory,” Extended abstracts of CHI'2002, conference on Human factors in computing systems. ACM, New York, NY, USA, 806-807, 2002, doi:10.1145/506443.506607.

[13] H. Jetter, S. Leifert, J. Gerken, S. Schubert, and H. Reiterer, “Does (multi-)touch aid users' spatial memory and navigation in 'panning' and in 'zooming & panning' UIs?,” Proceedings of AVI’2012, the International Working Conference on Advanced Visual Interfaces, Genny Tortora, Stefano Levialdi, and Maurizio Tucci (Eds.). ACM, New York, NY, USA, 83-90, 2012, doi:10.1145/2254556.2254575.

[14] ODICIS Web site https://odicis.org

[15] D. Z. Mangion, L. Bécouarn, M. Fabbri and J. Bader, “A Single Interactive Display Concept for Commercial and Business Jet Cockpits,” ATIO’2011, Conf. of the AIAA (American Institute of Aeronautics and Astronautics) on Aviation Technology, Integration, and Operations, 2011.

[16] A. Alapetite, R. Fogh, D. Zammit-Mangion, C. Zammit, I. Agius, M. Fabbri, M. Pregnolato and L. Becouarn, “Direct tactile manipulation of the flight plan in a modern aircraft cockpit,” Proceedings of HCI Aero’2012, International Conference on Human-Computer Interaction in Aerospace, Brussels, Belgium, 13 September 2012.

 ISBN: 978-1-61208-250-9

Effect of non-Unified Interaction Design of in-car Applications on Driving Performance, Situational Awareness and Task Performance

Julia Manner, Christopher Kohl, Michael Schermann, Helmut Krcmar
Chair for Information Systems
Technische Universität München
Boltzmannstraße 3
85748 Garching bei München, Germany
Email: {julia.manner, christopher.kohl, michael.schermann, krcmar}@in.tum.de

***Abstract*—It is common understanding that human-computer interaction (HCI) systems should be designed unified. However, ensuring a unified interaction design (UID) is a cost intensive and time-consuming venture. Especially the automotive industry struggles with exceeding costs and time-to-market pressure as drivers want to stay connected and informed while driving. Therefore, we investigated the effect of non-unified interaction design (NUID). We report on a simulator study with 44 participants in which we studied the effect of a NUID within an automotive HCI system consisting of five in-car applications. We measured the effect on driving performance, task performance and situational awareness when carrying out tasks. We found no significant effect of UIDs. We offer an explanation based on HCI and cognitive ergonomics literature.**

***Keywords*-interaction design; in-car applications; cognitive load; multi-tasking; multiple-tasks; task complexity**

I. Introduction

It is common understanding that HCI systems should be designed consistently to reduce the effort to use the technology [1]. Since the mid-nineties [2], literature considers consistent and standardized HCI design to be desirable. Accordingly, software designers generally implement UIDs. To give an example: the button ”x” at the top right of a window to close programs or the index tab with partial similar options like save have consistently been implemented. But, ensuring a UID for an IT system with various functionality is a cost intensive and time consuming venture [3], [4]. Every application not only needs to be tested on functionality and performance, but also on (subjective) characteristics like usability, design and compatibility which cannot be evaluated easily [4]. By contrast to this, the standard approach researchers acknowledge that usability needs to be engineered specifically for the context of use for the system under investigation [5]. In-car applications are now penetrating the automotive sector as it is widely recognized that drivers want to stay connected and informed while driving [6], [7]. Moreover, to serve this promising market, car manufacturers must provide various and innovative in-car applications [8], [9]. Consequently, the amount of money car manufacturers have to spend on providing in-car applications with a UID is exceeding [10]. We investigate the effect of a NUID for this new domain and – therefore - if the common understanding still holds in the automotive domain. We study the effect of a NUID of in car applications by proposing an experiment with two groups. The control group was provided with an in-car system that consists of five similarly designed applications with a UID. Therefore, the UID of a premium car manufacturer was imitated. The test group used the same applications but each application had a different interaction design. Besides, each group was instructed to carry out several tasks while driving in a simulator. A significant difference in consistency and standardization and hence the usability between the two groups was verified with a questionnaire after the usage including the empirically acknowledged construct of effort expectancy from the unified acceptance theory proposed by Venkatesh [1]. We measured the effect of NUID on driving performance, task performance, and situational awareness. No significant effect of NUID on driving performance or situational awareness was found. However, a significant effect on the driver’s task performance was measured. The remainder of this paper is organized as follows. In the next section, the theoretical background of our study will be explained. We then outline the study completely. Henceforth, we report on the results of our study and conclude with a summary of our findings and prospects of future research.

II. Background and theoretical framework

Literature of usability design [11] and on cognitive ergonomics [3], [12] emphasize the importance of ease of use. Thereby, standards and conformity are considered as crucial as they reduce the effort to operate HCI systems. In non-safety critical environments ease of operation is considered an important factor for user’s technology acceptance [13], but in a safety critical environment a NUID is endangering people [3]. Answering the need to measure the ease of use which is part of the acceptance, theory researcher proposed measuring instruments like the Technology Acceptance Model proposed by Davis in 1989 [13] or the Unified Theory of Acceptance and Use of Technology (UTAUT) model, which is a further development of the first [14]. Both models are empirically

 ISBN: 978-1-61208-250-9

validated and include a construct for measuring the effort users feel they need when operating the system. From this matured area of consumer behavior research, we use the construct of effort expectancy from the UTAUT model that evaluates the "degree of ease associated with the use of the IT system" and investigate if we provide a sound research design. Effort expectancy of users is measured by the following four items proposed by Venkatesh [15]:

1) Interaction with the system is clear and understandable
2) Ease of becoming skillful at using the system
3) Evaluation of the ease to use the system
4) Evaluation of the learnability of system use

Taking these in several domains [15], [16] applied indicators for effort expectancy into account, it is likely that a NUID increases the effort to use.

H1: A NUID leads to higher effort expectancy.

A. Driver's distraction by an additional task

Driving is a cognitively demanding multi-tasking activity [17], [18] which is competing for the limited resource of cognitive capacity [12] when an additional task like operating in-car applications or speaking on the phone take place simultaneously. Hence, it is not surprising that several studies show a negative effect of additional tasks on driving performance revealed by distraction [19], [20]. According to the American Automobile Association for Traffic Safety, a driver is distracted when he is "delayed in the recognition of information needed to safely accomplish the driving task because some event, activity, object or person within or outside the vehicle compelled or tended to induce the driver's shifting attention away from the driving task" [21]. A study by Lee et al. [22] even found out that by increasing the operating complexity of an IT system, likewise an increase of cognitive resources of the operator needed is recorded. Similar results were reported by Gkikas and Richardson [23] as they found out that an increasingly demanding conversation has a negative impact on driving performance. Moreover, Besnard and Cacitti [3] discovered that interface changes in a working environment cause accidents. Going along with these results, it can be concluded that an IT system which causes more effort to operate has a negative effect on driving performance [12]. Following this argumentation, we expect a NUID to have a negative effect on driving performance. This leads us to the following hypothesis:

H2: A NUID leads to a decreased driving performance

B. Task performance

Task performance is defined by the time a user needs to complete a task and the quality the user has performed [24]. Task quality can be measured by the number of mistakes made. According to Burns et. al. [25], a driver needs to maintain an overview of his overall plan. Task duration is a measure that needs to be monitored when studying such an effect. This is because an additional task competes with the driving task for limited cognitive resources. A negative effect on task performance can be assumed when the effort of carrying out the task is significantly higher through a NUID. Hence, we propose the following hypothesis:

H3: A NUID leads to a lower task performance

C. Situational Awareness

Proactive safe driving without considering additional tasks can be considered as multi-tasking [26]. Not only must the driver drive, but he also needs to monitor and process the environment. A concept that integrates this aspect is called "situational awareness". Endsley [27] simply defines it as "knowing what is going on around you" or in more detail he [28] defines it as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". The concept was introduced in aviation psychology more than 24 years ago. Therefore, in this study, not only the driving performance was measured, but also the situational awareness. A poor system design, which disregards usability guidelines such as unified interaction and therefore demands more effort to carry out an instruction, leads to a limited situational awareness [29]. Ma and Kaber (2005) found a similar result. They were able to prove that reducing the effort to drive by an adaptive cruise control leads to an improvement of situational awareness. For this reason, we claim that the working memory of an individual is limited [12], [19] which is in accordance to Baumann et al. [19] who state that cognitive load "withdraws resources necessary for comprehension of the current situation" which often leads to a phenomena called "looked but did not see". Therefore, we propose the following hypothesis:

H4: A NUID leads to a lower situational awareness

III. User Study

A. Experiment design

The study participants were undergraduate and graduate students from Technische Universität München with a mean age of 26 years. Each volunteer with a driver license was eligible to participate in the study. The study was announced in lectures as well as in flyers around the campus. We attracted participants by raffling an Apple iPhone 4S. In order to test the hypotheses, we divided the participants into two groups. Each group had to operate the same five in car applications to complete the same three tasks but with a different interaction design approach. In the control group, each application displayed the interaction design of a major car manufacturer. In the test group, the applications had different interaction designs. The experiment was conducted as follows. First, some general information about in-car applications and an agenda of the experiment was given to the participants. Then, each participant filled a questionnaire containing questions about demographics and driving experience. After completing the survey, the participants were shown how to drive the driving simulator and how to operate the given applications. After completing all tasks, a second questionnaire had to be filled out which contained questions regarding the tasks. These questions were applied to measure the situational awareness.

 ISBN: 978-1-61208-250-9

Figure 1. Driving Simulator

The answers were graded from 1 (wrong answer) to 3 (correct answer) because the participants could answer the questions in their own words. We also asked questions regarding the effort expectancy. Moreover, in this experiment, the participants were advised that the odds of winning the iPhone increase according to their driving performance. This incentive was used for substituting the natural human avoidance of suffering damage, which we would otherwise not be able to attain realistically with a simulator study. Furthermore, a constant speed of 40 km/h was required. All participants obeyed this request. Thus, it can be determined that throughout the whole experiment the complexity of the driving task is constant between the groups. The similarity of the tasks the participant had to perform and understand is ensured by an instruction and procedure document for the experiment supervisor. The instructions were read to all participants.

B. Driving simulator

To measure the driver's distraction, the study participants had to sit in a driving simulator as shown in Figure 1. They were instructed to fasten the seat belt and to use the provided Logitech Driving Force GT steering wheel and pedals. While the driving simulation was shown on a 27-inch screen, the applications were displayed on a separate screen. Both groups operated the in-car applications with the same controller with a jog dial to eliminate potential confounding influences from the controller [30]. The controller allowed the users to rotate and push the jog dial or nudge it in a specific direction. The controller also provides direct access to the functions "menu", "back", and "navigation" when enabled in the active application.

C. Application development

We used three different interaction designs among five in car applications. For example, one interaction design is to rotate the jog dial clockwise to select the next item. The items were always arranged vertically with the first item at the top of the screen. The applications were implemented with the SDK of a major car manufacturer to ensure the same design and functionality for both groups. We specifically excluded any signs of the manufacturer's brand to avoid any brand-specific prejudices or preferences. The applications were designed to help the participants to go through a predefined scenario consisting of three tasks which should make the experiment as realistic as possible. We also developed a logging application which recorded specific decisions the participants could take in the scenario while using the applications. The study participants were instructed to memorize as much of the displayed information as possible. Each participant operated a movie rental service before starting the actual experiment. This was to ensure that all participants were familiar with the controller and the general concept of using applications in a car-like environment. The first task in the scenario was to book a hotel room in a given hotel. The application displayed the most important information like the price for a room, a rating (1 to 5 stars) of the hotel and a short description which required the driver to look on the screen for a longer time span in order to get all necessary information and complete the booking process. After having completed the booking process, the hotel had to be set as the new navigation target and the route had to be displayed on a map in the navigation application. The second task was to book a table for two people in a restaurant near a tourist attraction in Munich at 9 pm. The restaurant then had to be added to the navigation targets and a map with the current route had to be displayed on a map in the navigation application. Again, many details were provided in the restaurant application, out of which some had to be remembered. The last task was to display the weather for the evening. The weather application displays the weather forecast with large symbols so that a short glance at the screen was sufficient to get the information.

TABLE I
DEFINITION OF MDEV

Description	Definition
Number of measured data points	N
Distance difference	$\Delta y_i = \frac{y_{i+1} - y_{i-1}}{2}$ (1)
Actual-Reference-Deviation	$x_{a,i} = \lvert x_i - x_{i,ref} \rvert$ (2)
Length of trip	$S = \sum_{i=1}^{N} \Delta y_i$ (3)
Mean Reference-Actual Deviation (MDEV)	$\overline{x_a} = \frac{1}{S} \sum_{N}^{i=1} x_{a,i} \Delta y_i$ (4)

D. Measurement instruments

The lane change test (LCT) is acknowledged by the International Organization for Standardization for being cost-efficient, reliable and a simple tool to measure driver's performance while carrying out an additional task [31], [32]. The LCT

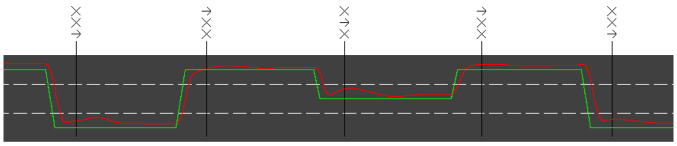


Figure 2. Driving Simulator

measures the mean reference-actual deviation (MDEV), i.e. the mean deviation between the reference lane and the actual driven lane. The formal definition of MDEV is shown in Table I. Figure 2 shows a graphical representation of a section of a LCT file. The test was applied by several authors which acknowledged the suitability of the tool [25], [32]. Bruyas et al. [33] reviewed the sensitivity of the LCT and found it excellent for differentiating between the execution of an additive task and the sole driving task. The LCT also allows setting a specific maximum speed and enables to provide the same driving challenge for each participant so that a direct comparison is possible and thus makes it especially suitable for experimental studies with simulators [31]. Hence, we chose the LCT for measuring the driver's distraction in our study. The LCT tool was configured to allow a maximum speed of 40 km/h in order to simulate a typical driving situation in Munich with an average speed of 32 km/h [34]. The study participants were instructed to accelerate fully in order to maintain the same speed throughout the driving part of the experiment. We incentivized serious driving by including the driving performance as a criterion for winning the iPhone. The LCT was started after each participant indicated that he or she was ready.

IV. Data analysis

A. Sample characteristics

44 licensed drivers (36 male and 8 female) with a mean age of 26 participated in this study. This particular age range was selected to represent the generation of future drivers. The mean mileage was 5,200 kilometers per year. In the study, there were 30 smartphone user and ten participants had experience with in-car applications. In each group there were 22 participants. Although we assigned each participant with a random distribution, there were some differences. In the NUID group the amount of females was slightly higher (5) than in the group with the UID (3). Furthermore, the mean mileage per year in the group with UID was higher (12.368km) than in the group with NUID (8.840km).

B. Statistical analysis

We first inspected the boxplots of the data, if there were any outliers. We found outliers in the measurements of MDEV, task completion time and effort expectancy (see Figure 3-6). We checked that we did not make an error and therefore chose to keep the outliers. As the distributions of these values have the same shape in both groups, we applied a Mann-Whitney test to determine, if there are differences between the groups.

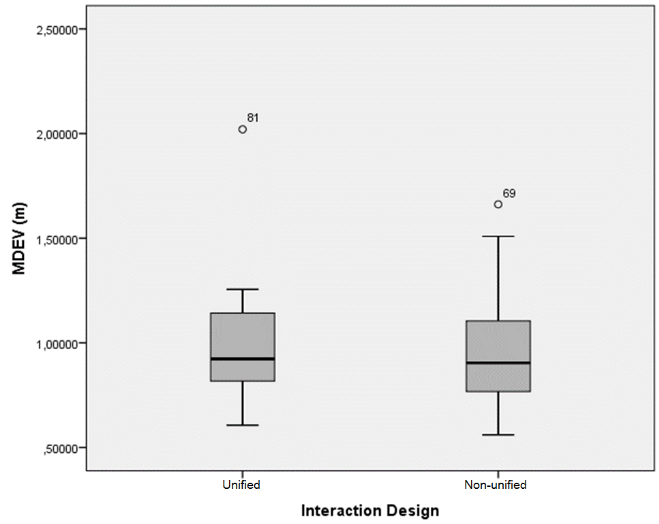


Figure 3. Boxplot of the MDEV distribution

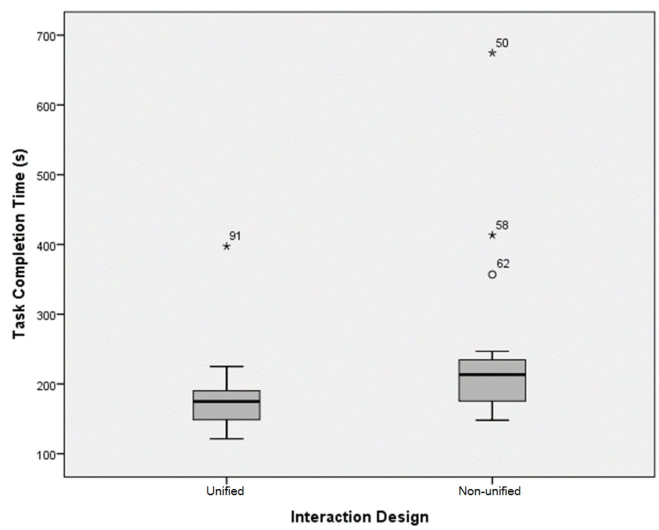


Figure 4. Boxplot of the task completion time distribution

MDEV: The test did not show significant differences in MDEV between the UID (Mdn = 0.923m) and the NUID (Mdn = 0.904m), U = 221, z = -0.493, p = 0.662. The NUID group even has a slightly lower MDEV (see Figure 3) i.e. the participants in this group drove slightly better than in the UID group.

Task performance: Task performance is measured by the quality of execution and the failures occurring by carrying out the tasks., e.g., booking a table only for one person instead of two. The test shows a significant difference between the UID (Mdn = 175s) and the NUID (Mdn = 213s), U = 363, z = 2.840, p = 0.005. Therefore, the UID group completed the tasks significantly faster. Figure 4 shows that there are also more outliers in the non UID group. The circle indicates a normal outlier, that has a distance of 1.5 box lengths to the edge of their box whereas the star indicates an extreme outlier,

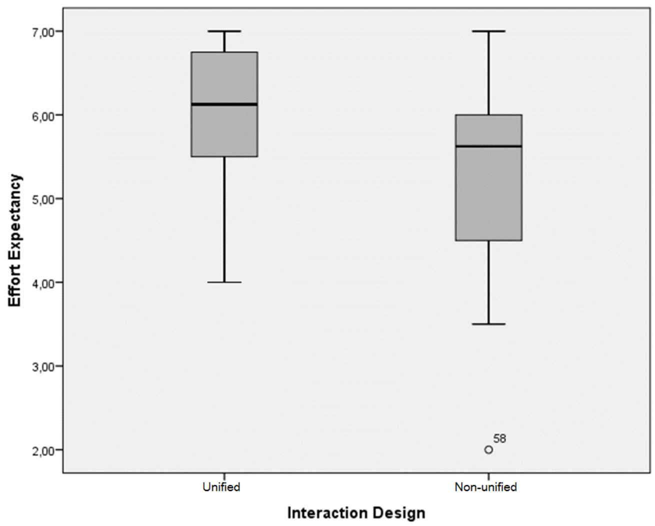


Figure 5. Boxplot of the effort expectancy distribution

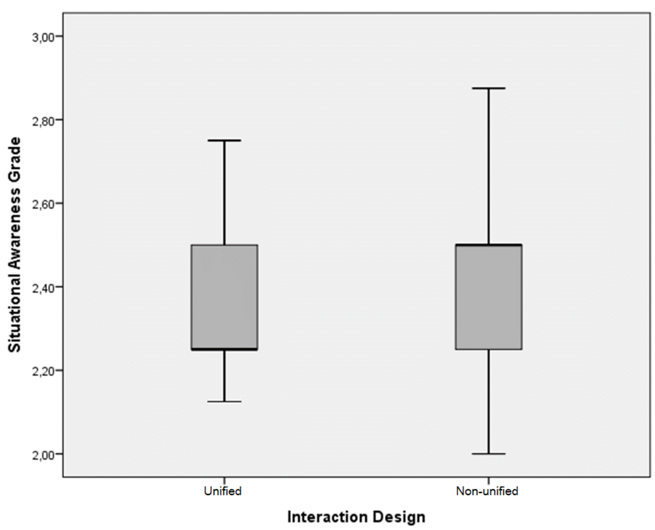


Figure 6. Boxplot of the situational awareness grades distribution

that has a distance of 3 box lengths to the edge of their box. We assume that the outliers had serious difficulties to operate the applications. The second aspect of task performance, the quality of task completion, showed no difference at all. In both groups, we had two participants who deviated from the given scenario in one aspect by e.g. booking a room in the wrong hotel. We therefore only consider the task completion time for measuring task performance.

Effort expectancy: The test shows significant differences in the mean of all four dimensions of effort expectancy (see theoretical background) between the UID (Mdn = 6.13) and the NUID (Mdn = 5.63), U = 158, z = -1.982, p = 0.047. So the participants in the NUID group recognized a decrease of the ease of use as they have rated it significantly lower than the participants in the UID group (see Figure 5).

Situational awareness: In order to measure situational awareness, the participants had to answer several questions about the characteristics of the hotel they had booked or the restaurant where they freely chose to book a table. The answers to the questions were graded and we examined the mean grade of all questions. There were no outliers in the boxplot of the mean grades (see Figure 6), the Q-Q plot showed a normal distribution and Levene's test for equality of variances showed the homogeneity of variances for the mean grades (p = 0.657). Therefore, we applied an independent samples t-test. The t-test did not show a significant difference between the UID (M = 2.39, SD = 0.200) and the NUID (M = 2.43, SD = 0.195); t(42) = -0.667, p = 0.508 . Both groups showed good results regarding situational awareness as the grade 3 was the best possible grade if every answer was answered correctly.

C. Result discussion

Our proposed hypotheses about the effect caused by a NUID in the automotive domain, were derived from the fields of situational awareness, marketing and previous studies of driver's distraction and cognitive ergonomics. We could confirm our first hypothesis. Indeed, a NUID leads to a significantly higher effort expectancy perceived by drivers. Contrary to that, we could not confirm H2 and H3. There was neither a significant difference in driving performance nor in situational awareness detected by the applied statistical tests. H4 however was confirmed. The task performance is reduced by a NUID. Although there was no difference in task quality, the second measure of task performance the time to complete the tasks was measured to be significantly longer. According to Burns et al. [25], task completion time is only an indirect measure that does not have an impact on driver's safety. Literature is consistent that the cognitive capacity of humans is limited [35]. When humans perform multiple tasks simultaneously, these tasks are in competition to each other for the limited resource of cognitive capacity. Based on Besnard and Cacitti [3], we notice that humans allocate that resource according to their respective goals. Based on our results, we claim that performing a task with a NUID causes not only more effort but also a different cognitive load for the participants. Complexity is an acknowledged increase factor for cognitive capacity [20]. Hence, we conclude a NUID that leads to a higher number of user interactions for successful selection of bookings or categories. This not only enhances complexity but also cognitive load. This conclusion is also in agreement to Bensard and Cacitty [3] who report the changing interface design increases the cognitive resources needed for operating a task. Our observations of the effect of an additive task cognitive load increase are in harmony with previous research and suggest that drivers allocate their existing limited cognitive capacity proactively according to their goals [3]. We found no evidence that drivers change the allocation of cognitive resources for driving and situational awareness while driving, if the task complexity of the additive task is increased. Our results rather show that drivers start to multiplex the additive task. Based on the assumption of a limited cognitive capacity, we conclude that this is the only option for drivers to carry out the task without enhanced failure rate or a performance decrease in driving or situational awareness. Moreover, we assume a connection between cognitive load intensification and task completion time in a multi-task situation where the additive task has a low priority for the user as shown in Figure 4. We emphasize that this correlation is only appliciable for additive low priority tasks. In safety critical evironments where multi-tasking is necessary, the allocation decision of the user may not be possible as presented [3]. Results observing from other authors like Ma and Kaber [36] which study the effect of increasing conversations complexity while driving reveal opposing results. However, we argue that a driver needs to answer his conversation partner. That is why a conversational task while driving cannot be easily multiplexed. This assumption is also in accordance with previous results. According to Broadbent [37] and Salvucci et al. [17], the concept of multi-tasking is always carried out by a series of temporal multiplexing. People are multiplexing when they are "sequentially allocating the available attention on a task" [17]. Our results indicate that the time slots allocated for situational awareness, driving and accomplishing tasks remain the same as the respective goals of the participants do not change [3]. Thus, the attention that is left to allocate for the tasks remain the same, users need more time when the task is more complex. The presented discussion shows that although our results may surprise, they can be integrated very well in the previous research and complement it with new insights. However, there are some limitations to our study. First, we examined the effect of ignoring standard and conformity only on one example: the UID. Although this is an outstandingly important example [11], there are different manipulations possible whose effects could be completely different in the automotive domain. Second, we provided a sound research design to achieve the presented expected results. The two treatments presented were demonstrated to cause significantly different effort expectancy. Although effort expectancy is an empirically tested construct that is widely acknowledged measure, it is not a measurement instrument to measure cognitive load. An experimental design also has a severe drawback. Building an artificial environment like proposed does not reflect the real world. Especially, events while driving are difficult to simulate due to many unpredictable factors. Therefore, the external validity of our experiment design is only partly fulfilled [38]. Furthermore, although the LCT is an ISO standard [31], it only simulates driving to a certain degree. For this experiment, we propose a speed of 40 km/h. Thereby, we were simulating the average speed of driving in a city like Munich [34]. Due to this fact, the results may not be generalized for other driving situations. At last, this study is limited by a gender bias as we recruited more male than female students. According to Petzold et al. [39] the comparison of simulator studies shows a difference between driving and task performance regarding the gender.

V. Conclusion and Future Work

We investigated the effect of a NUID of in-car applications on drivers. As driving solely can be considered as multitasking [18] since the driver must not only drive, but also realize and comprehend the situation around him, we measured the driving performance as well as the situational awareness. Besides, the task performance was measured to cover also this factor of cognitive capacity consumer. Surprisingly, we could not find a significant effect on driver's performance and situational awareness. Only the task performance decreased. Although, the task failure rate showed no significant effect between the groups, task completion time of the drivers increased significantly. According to Burns et al. [25], task completion time is only an indirect measure that does not have an impact on driver's safety.

The proposed results also contribute to literature. Although situational awareness and cognitive capacity are seen as important concepts for understanding driving performance, they are not well understood [35], [36]. There are key gaps in literature as researchers still wonder what compromises situational awareness while driving [35]. Although, this research does not provide factors compromising situational awareness, it shows that an IT system that can be suspended and respectively can be operated slower have not a significant negative factor on drivers. Moreover, it indicates that there is a correlation between task complexity and therefore cognitive demand and task completion time for additive tasks in the automotive domain.

Our results also have a practical value. We could not find evidence of an endangering of a NUID. Consequently, this results leads in-car application designer to further test the topic. As the costs of ensuring UID are enormous, a lesson of some tests for usability could be a promising perspective. Besides, the experiment shows that the adoption of results from other domains is critical and some guidelines for design should be tested again in this domain.

Finally, we suggest two further research directions. According to Endsley [12] and others [40] the cognitive load needed for operating a system can be reduced by experience and further enhance performance on the primary driving task. Repeating the experiment with participants that are trained to operate the in-car applications could provide further insights. Furthermore, the effect of predefined pauses when operating the system while driving, forcing more attention to the road should be investigated. As the allocation of attention for multiple tasks is a key sub-skill of situational awareness [26] new insights for road safety could be detected.

References

[1] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, 2003.

[2] J. Nielsen, L. A. Blatt, J. Bradford, and P. Brooks, "Usability inspection methods," p. 2, 1994.

[3] D. Besnard and L. Cacitti, "Interface changes causing accidents. an empirical study of negative transfer," *Human-Computer Studies*, vol. 62, p. 20, 2005.

[4] K. Ehrlich and J. Rohn, "Cost justification of usability engineering: A vendor's perspective," pp. 73–110, 1994.

[5] C. Harvey, N. A. Stanton, C. A. Pickering, M. McDonald, and P. Zheng, "A usability evaluation toolkit for in-vehicle information systems (iviss)," *Applied ergonomics*, vol. 42, no. 4, p. 12, 2011.

[6] F. Alt, D. Kern, F. Schulte, B. Pfleging, A. S. Shirazi, and A. Schmidt, "Enabling micro-entertainment in vehicles based on context information," p. 8, 2010.

[7] D. Basacik, N. Reed, and R. Robbins, "Smartphone use while driving - a simulator study," Transport Research Laboratory, Tech. Rep., 2012.

[8] Mercer, "Future automotive lndustry structure (fast) 2015," Heinrich Druck + Medien, Tech. Rep., 2004.

[9] R. Reichwald, H. Krcmar, and S. Reindl, *Mobile Dienste im Auto der Zukunft Konzeption, Entwicklung, Pilotierung*. Lohmar-Köln: Reichwald, R., Krcmar, H., Reindl, S., 2007.

[10] A. Zauner, H. Hoffmann, J. M. Leimeister, and H. Krcmar, *Automotive Software and Service Engineering (ASSE) - an exploration of challenges and trends in an industry experts' view*. Norderstedt: Books on Demand, 2010, pp. 61–74.

[11] J. Nielsen, *Usability engineering*, 1st ed. Boston: Academic Press, 1993.

[12] M. R. Endsley, B. Bolté, and D. G. Jones, *Designing for Situation Awareness: An Approach to User-Centered Design*. New York: Taylor & Francis Inc., 2003.

[13] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User acceptance of computer technology: A comparison of two theoretical models," *Management Science*, vol. 35, no. 8, pp. 982–1003, 1989.

[14] V. Venkatesh, V. Ramesh, and A. P. Massey, "Understanding usability in mobile commerce," *Communications of the ACM - Mobile computing opportunities and challenges*, vol. 46, no. 12, pp. 53–56, 2003.

[15] V. Venkatesh and V. Ramesh, "Web and wireless site usability: Understanding differences and modeling use," *MIS Quarterly*, vol. 30, no. 1, pp. 181–205, 2006.

[16] M. Williams, N. Rana, Y. Dwivedi, and B. Lal, "Is utaut really used or just cited for the sake of it? a systematic review of citation of utaut's originating article," pp. 1–13, 2011.

[17] D. Salvucci, E. Boer, and A. Liu, "Toward an integrated model of driver behavior in cognitive architecture," *Transportation Research Record*, vol. 1779, no. 1, p. 7, 2001.

[18] A. Sethumadhavan, "In-vehicle technologies and driver distraction," *Ergonomics in Design: The Quarterly of Human Factors Applications*, vol. 19, no. 4, p. 3, 2011.

[19] M. Baumann, T. Petzoldt, C. Groenewoud, J. Hogema, and J. Krems, "The effect of cognitive tasks on predicting events in traffic," p. 11, 2008.

[20] K. Young, M. Regan, and M. Hammer, "Driver distraction: A review of the literature," Monash University Accident Research Centre, Tech. Rep., 2003.

[21] J. R. Treat, *A Study of Precrash Factors Involved in Traffic Accidents*. Ann Arbor, MI: University of Michigan Highway Safety Research Institute, 1980.

[22] J. D. Lee, B. Caven, S. Haake, and T. L. Brown, "Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 43, no. 4, pp. 631–640, 2001.

[23] N. Gkikas and J. Richardson, "Behaviour & information technology the impact of verbal interaction on driver lateral control : an experimental assessment," *Behaviour & Information Technology*, vol. 31, no. 6, p. 12, 2012.

[24] A. Barón and P. Green, "Safety and usability of speech interfaces for in-vehicle tasks while driving," University of Michigan, Transportation Research Institute, Tech. Rep., 2006.

[25] P. C. Burns, P. L. Trbovich, T. McCurdie, and J. L. Harbluk, "Measuring distraction: Task duration and the lane-change test (lct)," p. 3, 2005.

[26] L. L. Gugerty, *Situation awareness in driving*. Boca Raton, FL: CRC Press, 2011, ch. 26, p. 25.

[27] M. R. Endsley, *Theoretical underpinnings of situation awareness: A critical review*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 2000, pp. 1–24.

[28] M. Endsley, "Design and evaluation for situation awareness enhancement," in *Human Factors Society Annual Meeting*. Human Factors Society, 1988, pp. 97–101.

[29] C. Bolstad, A. Costello, and M. Endsley, "Bad situation awareness designs: what went wrong and why," in *Proceedings of the 16th international ergonomics association world congress*. Elsevier Science, 2006.

[30] K. Bengler, M. Herrler, and H. Künzner, "Usability engineering accompanying the development of idrive," *Informationstechnik und Technische Informatik*, vol. 44, no. 3, pp. 145–152, 2002.

[31] S. Mattes and A. Heallén, *Surrogate distraction measurement techniques: The lane change test*. Florida: CRC Press, 2009.

[32] J. L. Harbluk, J. S. Mitroi, and P. C. Burns, "Three navigation systems with three tasks: Using the lane-change test (lct) to assess distraction demand," pp. 24–30, 2009.

[33] M.-P. Bruyas, C. Brusque, H. Tattegrain, A. Auriault, I. Aillerie, and M. Duraz, "Consistency and sensitivity of lane change test according to driving simulator characteristics," *IEET Intelligent Transport Systems*, vol. 2, no. 4, p. 306–314, 2008.

[34] P. Olson and K. Nolan, "Europe's most congested cities," 2008.

[35] P. M. Salmon, N. A. Stanton, and K. Young, "Situation awareness on the road: review, theoretical and methodological issues, and future directions," *Theoretical Issues in Ergonomics Science*, vol. 13, no. 4, p. 22, 2012.

[36] R. Ma and D. B. Kaber, "Situation awareness and workload in driving while using adaptive cruise control and a cell phone," *International Journal of Industrial Ergonomics*, vol. 35, pp. 939–953, 2005.

[37] D. E. Broadbent, *Perception and communication*. London: Pergamon Press, 1958.

[38] L. Nilsson, *Contributions and limitations of simulator studies to driver behavior research*. London, Washington DC: Taylor & Francis, 1993, ch. 401-407.

[39] T. Petzoldt, N. Baer, and J. Krems, "Gender effects on lane change test (lct) performance," pp. 90–96, 2009.

[40] K. S. O'Brien and D. O'Hare, "Situational awareness ability and cognitive skills training in a complex real-world task," *Ergonomics*, vol. 50, no. 7, p. 27, 2007.

 ISBN: 978-1-61208-250-9

Studying Depth in a 3D User Interface by a Paper Prototype as a Part of the Mixed Methods Evaluation Procedure

Early Phase User Experience Study

Leena Arhippainen, Minna Pakanen, Seamus Hickey
Intel and Nokia Joint Innovation Center/ Center for Internet Excellence
P.O. Box 1001, FI-90014 University of Oulu, Finland
leena.arhippainen@cie.fi, minna.pakanen@cie.fi, seamus.hickey@cie.fi

***Abstract*— A principal characteristic of three dimensional user interfaces is that it contains information in the 3rd axis. Visually, this information is presented as being placed further away from the screen or having depth. A consequence of this is that information can be occluded. Determining the optimal amount of depth levels for specifically sized icons is important in the design of 3D user interfaces. This paper investigates the depth placement of objects of a three dimensional user interface on a tablet device at the early stage of the development process. We present mixed methods evaluation with a paper prototype with a focus on the users' subjective experiences. Users were presented with concepts of different depth levels, with and without 3D objects. The findings indicate that users' preference was for depth levels 3-5. We recommend designing 3D UIs with a controllable depth by starting with a few depth levels and increasing them automatically based on the amount of 3D objects. Also, it is important to give a user a possibility to customize depth levels when needed. This paper provides user preference information on depth for 3D UI designers and developers, especially in the context of a touch screen tablet device.**

Keywords-3D UI; depth; touch screen tablet; paper prototype; user experience.

I. INTRODUCTION

Three dimensional (3D) graphical user interfaces (GUIs) have been studied for many decades and they are still actively researched [1][2][7][10][14]. Typically, interaction with 3D user interfaces involves dealing with information and objects that are spatially organized in three dimensional virtual space. The number of objects, their size and issues of occlusion must be defined and evaluated with users in order to provide a positive user experience. To show large amounts of information, 3D objects are spatially organized at different levels of depth from the 3D camera, or what we call depth levels. 3D research and development is a large and diverse area. Studies exist about appropriate 3D UI metaphors and depth for touch screens in PC environments [1][11]. However, there is not a clear answer for what kind of 3D UI and depth users would actually prefer for touch screen tablet devices.

Cipiloglu et al. [5] present a framework for enhancing depth perception in computer graphics. Different depth cues help users to perceive the spatial relationships between the objects. Earlier studies indicate that spatial organization of information enables efficient access to objects in graphical user interfaces. Cockburn and McKenzie [6] studied the effectiveness of spatial memory in 2D/3D physical and virtual environments and compared 2D, 2.5D and 3D interfaces. However, this experiment was conducted with a PC and mouse and one would expect interaction to be different than with a touch screen tablet device. One influencing factor for a user-tablet interaction with touch screen is the size of the objects, widgets and icons. According to Budiu and Nielsen [3], the target size for 2D UI widgets is 1*1cm for touch devices. In a 3D UI on a tablet device, there are several aspects that can have an influence on how users perceive the space and how they are able to interact with 3D objects in depth through touch screen. This issue has not been studied much in a mobile tablet device context. In a hybrid 2D/3D UI study, Salo et al. [15] found that a large amount of 2D overlay icons decreases the interaction with 3D objects which are embedded in a 3D virtual environment.

This paper presents how we used a paper prototype as a part of mixed methods evaluation procedure at the early phase of the design process to find out the optimal depth levels in 3D graphical user interfaces. Later, we developed a virtual prototype as well in order to see how users' preferences for depth in a 3D UI compare with the results from the paper prototype. Based on the user evaluations with both paper and virtual prototypes, we propose that the depth levels 3-5 could be the most preferable depth for the 3D UI as a default starting point, depending on the system context.

II. METHOD OF STUDY

Paper prototyping is a widely used method in the human-computer interaction (HCI) field, especially in the user-centered design (UCD) process [16]. The aim of the UCD process is to support developing systems that meet the user's expectations and needs. ISO 13407 [8] defines a prototype as: "*representation of all or part of a product or system that, although limited in some way, can be used for evaluation*". According to the Buchenau and Fulton Suri [4], prototypes are "*representations of a design made before final artifacts exist and they are created to inform both design process and design decisions*". Prototypes can range from sketches to different kinds of models, which depict the design as follows

 ISBN: 978-1-61208-250-9

"looks like," "behaves like," "works like" [4]. One benefit of using prototypes is that they can facilitate exploring and communicating propositions about the design and its context.

The paper prototyping itself as a method is not a new idea. Typically, it has been used for improving usability of the UI. Our focus was on subjective user's experiences. When studying user experiences in the early design phase, it is important to use suitable research methods. Although the interest in UX in industry and academy has been high over a decade, there are not enough systematic methods regarding how to evaluate user experiences [17]. Especially, there is a need to develop and use low-cost methods for UX evaluation and utilize the collected information in the early phase of the design and development processes [18]. ISO 9241-110 [9] defines user experience as: "*a person's perceptions and responses that results from the use and/or anticipated use of a product, system or service"*. Therefore, user experiences should be evaluated before, during and after the use [17].

A. Concept Design

In the early stage of our research, we did a concept design phase to explore different ideas for the visual design of the 3D UI. In this phase, we drew approximately 100 sketches of different 3D UIs for touch screen tablet devices. From those sketches we selected the most relevant examples for the further design. Among those sketches there were two sketches about 3D UI utilizing depth. The first example illustrates the 3D objects on top of the VE (Fig. 1a) and the second the same example in a customization mode where the grid is visible and background VE is invisible (Fig. 1b). Based on these sketches we created a paper prototype in order to study depth levels for 3D UIs on tablets.

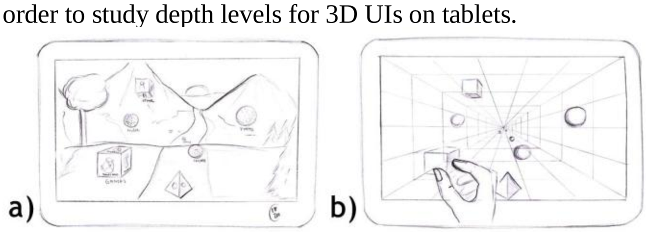


Figure 1. The sketches of 3D UI and objects on A) a virtual environment and B) with a grid background.

B. Evaluation Procedure

We developed a paper prototype in order to study users' preferences for depth levels in 3D UIs on tablets in the early design phase. We used our depth level paper prototype as a part of mixed methods evaluation procedure where we illustrated ten 3D UI concepts to the users by using various types of prototypes. Fig. 2 presents the contents of the whole evaluation procedure. This paper focuses only on the findings gathered from the phase 5: 2) selection tasks A-C (Fig. 2). It was important to introduce participants to the 3D UI topic and show different examples in the evaluation phases 1-4 before the depth level evaluation (Fig. 2). Findings from the phase 1 are reported in [12] and from the phases 2 and 3 in [13].

Evaluation Procedure:

1. 2D/3D icon comparison tasks (tablet prototypes)
2. Four 3D UI Concept evaluations (tablet prototype)
3. 3D UI Use case evaluation tasks (tablet, PC, paper prototypes)
4. Contact and Square UI evaluation (tablet prototype)
5. 1) 3D UI space test (paper prototype)
   Space form selection for 3D UI (examples A1-H8)

   2) 3D UI depth level selection tasks A-C (paper prototype)
   Select the level 1, 2, 3, 4, 5, 10, 15 or ∞ based on which one you prefer or you think you could control:
   A. Selection Task: Without objects
   B. Selection Task: With ordered 3D objects
   C. Selection Task: With unordered 3D objects
6. 3D UI concept evaluations (PC prototype, video)
7. Self-expression tasks (drawing template).

Figure 2. The 3D UI Concept evaluation procedure with mixed methods and the depth level paper prototype evaluation in the phase 5. 2) A-C.

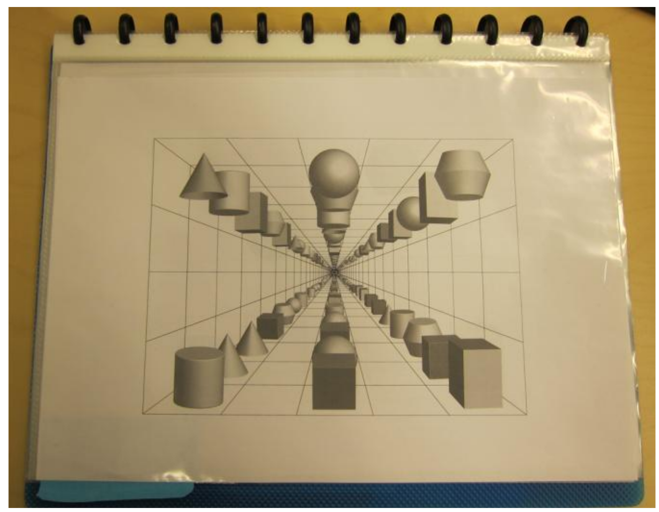


Figure 3. An example page of the paper prototype (Infinity depth).

C. Depth Level Paper Prototype and Selection Tasks

The paper prototype was created in such a way that a UI example would be comparable with commercial off-the-shelf touch screen tablet devices with 2D icons in a 2D UI (e.g., Apple iPad and Samsung Galaxy Tab). For instance, in these kinds of tablets, there are 4*5 application icons presented per screen. Therefore, our grid example included 4*5 icon areas as well. The depth in each level is the space that a 3D icon of an application requires. Fig. 3 illustrates one example page of our paper prototype. The size of the grid was 23.8 centimeters (almost equal to the iPad's screen size, which is 9.7 inches) on a size A4 white paper. Fig. 4 illustrates depth levels without objects (A), with ordered (B) and unordered 3D objects (C). In each selection task (A-C), we had eight depth levels: 1, 2, 3, 4, 5, 10, 15, infinity (∞) and users were asked to select one level in each task (A-C) based on which depth level they prefer (Fig. 2).

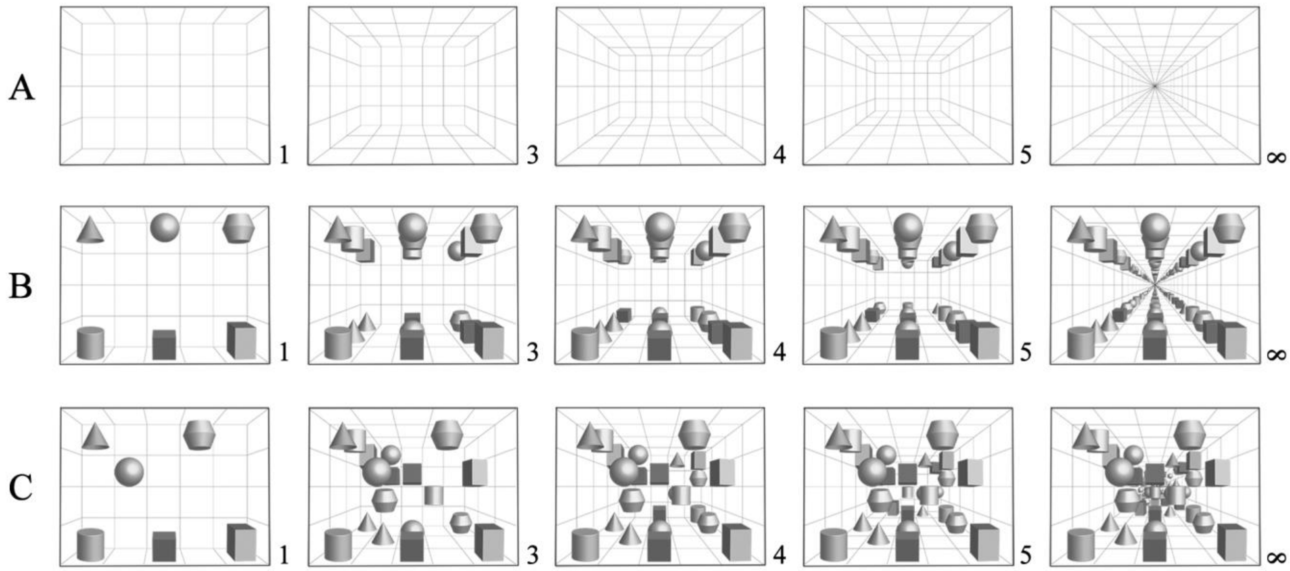


Figure 4. Example depth levels 1, 3, 4, 5 and ∞ without objects (A), with ordered (B) and unordered 3D objects (C).

We decided to create depth levels with a different set of 3D objects, because of the two reasons. First, in the 3D UIs, objects can be placed anywhere and they can be occluded. Second, in the tablet devices, a user can have a different amount of application icons and widgets.

D. Participants

The evaluation was conducted with 40 participants (15 female, 25 male), whose age varied from 23 to 52 years (averaging 35) (Fig. 5). Users had prior experience with touch screen devices, either tablets or phones.

III. FINDINGS

The following Subsections present which depth levels users preferred in the selection tasks A-C (without objects and with ordered and unordered 3D objects) in the 3D UI concept evaluation (Fig. 2). Users' subjective experiences are also cited.

A. Depth Selections without Objects

Without 3D objects, depth levels 3 (40%) and 4 (35%) were clearly the most selected choices (Fig. 6). According to the feedback, participants made their selections based on how many icons or applications they could place in the space and how they could select them by touching. For instance, depth level 1 was regarded as too plain or tight and only a few applications could be located on the periphery.

One person, who selected level 2, commented: "Here could be 36 icons on sides and ceiling". Participants understood that these icons would occupy the same volumetric space as a 2D icon. Subjects selected depth levels from 2 to 4; because they did not want the background grid to be too small. Participants understood that the background grid decreases when the depth increases. They thought the background area is meant for open applications, which is easily viewable. A person, who selected level 3, said that the depth depends on the physical size of a finger, the finger will "poke" many icons if the space has too much depth.

Participants who selected depth levels from 3 to 5 justified their selection by referring to memory: "*It would be impossible to remember where some objects are*". The comments on depth levels 10, 15 and infinity were: "*too deep*", "*too small periphery*", "*difficult to control*" and "*cannot use anymore*". One person commented that the infinity level could be a suitable solution for thousands of music files.

Figure 5. In the concept evaluation with the paper prototype, a participant is selecting the depth level 5 and commenting the 3D object selection.

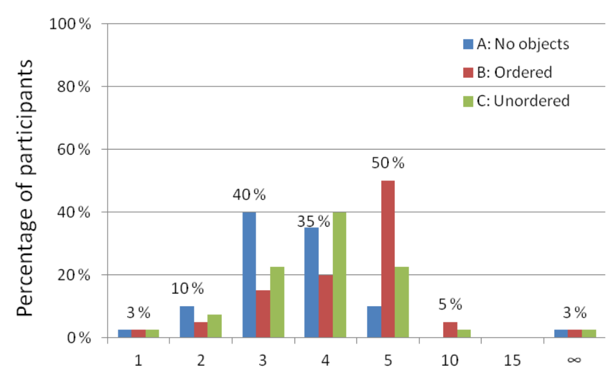


Figure 6. Participants' selections in the tasks A-C.

B. *Depth Selections with Ordered 3D Objects*

The comments from participants during the selection of the depth level without objects, revealed how they would place 2D icons on the surfaces. When participants saw depth levels with 3D objects, they perceived how those would appear and be located in the 3D space. When 3D objects were organized in the space, the depth selection was deeper, half of the participants selected depth level 5 (Fig. 6). Users, who selected levels 4 or 5, said that "*from these levels they can remember, control and select icons*". In Fig. 5, a participant is selecting the depth level 5 because in that level he can see and select also occluded objects. One person, who selected level 2, commented that "*If objects would be organized only on the sides, the depth level could be even deeper*".

In this task, people also counted how many objects there would be available for them, and thus, level 2 was regarded as too small. Depth level 1 was regarded as boring. One person, who selected level 5, commented that "*The level 1 shows that I'm poor, I don't have many things happening in my life*". She explained this comment by comparing how many applications she would need in her private and professional life. A person, who selected level 10, said that level 1 would be enough for his mother (e.g., elder and non-technically oriented people). A majority of the comments of depth levels 10, 15 and infinity were negative and related to issues like visual appearance, controllability and memory. One person said: "*It does not feel coequal, because some item is behind the others*". One user thought that depth could increase according to the amount of applications.

C. *Depth Selections with Unordered 3D Objects*

When 3D objects were not organized, 40% of all users selected the depth level 4 (Fig. 6). Participants, who selected level 4, said that they would like to be able to change their viewpoint or perspective to see behind the objects. Depth levels 1 and 2 were regarded as boring. Some users thought that level 2 causes claustrophobia. Depth levels 3-5 got comments, mainly relating to controllability. Comments on level 5 or more were mainly negative. Users wondered how visible and recognizable the icons would be. They thought that different 3D object shapes could make them recognizable. Some arguments reflect more on the users' personality, for instance, one person selected level 5 and said: "*I like a certain type of chaos*". Another person who selected level 10, said: "*I'm not a minimalist*".

IV. VALIDATION OF THE FINDINGS

This section presents how we later validated UX findings gathered by the paper prototype. We developed a virtual prototype on a tablet device and conducted the similar depth level evaluation with both the paper and virtual prototypes. However, this test did not include the whole 3D UI concept evaluation procedure (Fig. 2).

A. *Development of Virtual Prototype on a Tablet*

We evaluated our paper prototype method by developing a virtual prototype and conducting the same depth level evaluation with the virtual prototype. In order to be able to conduct a similar evaluation, we designed depth levels with the same visual appearance (Fig. 7). However, in the final virtual prototype, we implemented more depth levels, because we also wanted to study levels from 6 to 20. Another reason for small differences on the appearance was that the virtual prototype was developed by using a 3D program and its camera perspective. This made the background grid appear in a slightly different size in the virtual prototype than in the paper prototype (Fig. 8). Therefore, we measured the hypotenuse of the background grid at each depth level in both of the prototypes and compared them. The difference in the size of the hypotenuse was approximately +/- 1 centimeter. This small difference has been taken into account in the comparison of the results (Fig. 11b).

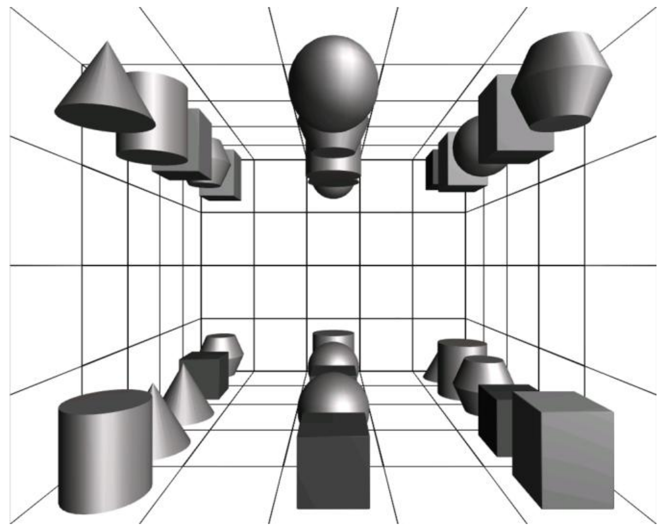


Figure 7. The screenshot of the 3D UI depth example in a virtual prototype (depth level 5).

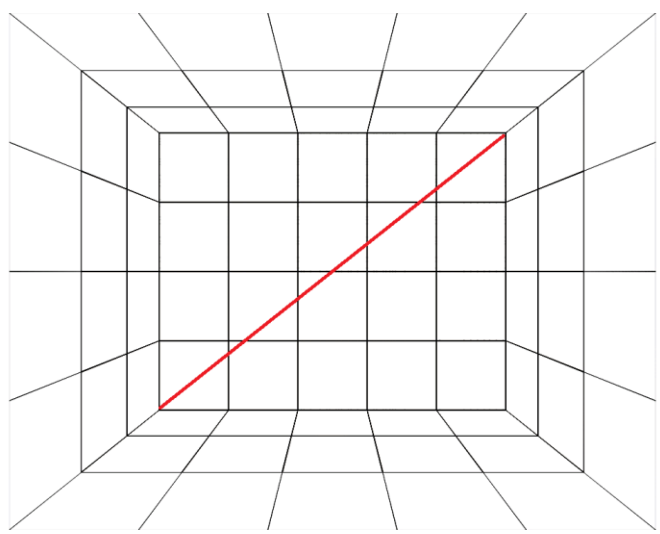


Figure 8. Size of the hypotenuse of the background grid slightly varied with some depth levels in the paper and virtual prototypes.

The virtual prototype was developed on a tablet device (9,7") and users were able to adjust the depth levels by pressing the screen continuously with one finger and then swiping forward (swipe up) and backward (swipe down) with another finger (Fig. 9). This interaction model was the main difference in comparison to the paper prototype, in which the depth levels were adjusted by turning pages of the binder (Fig. 10a).

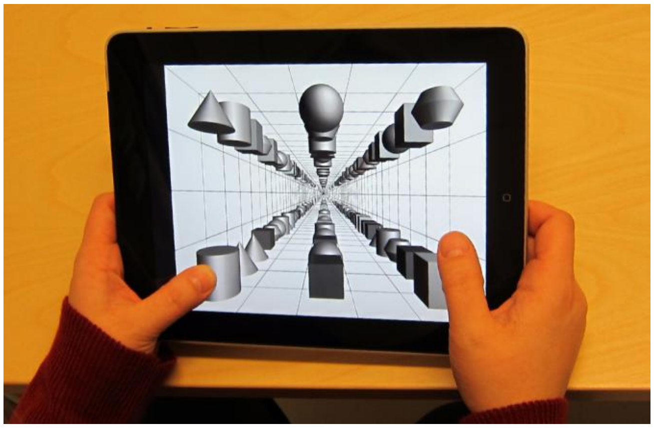
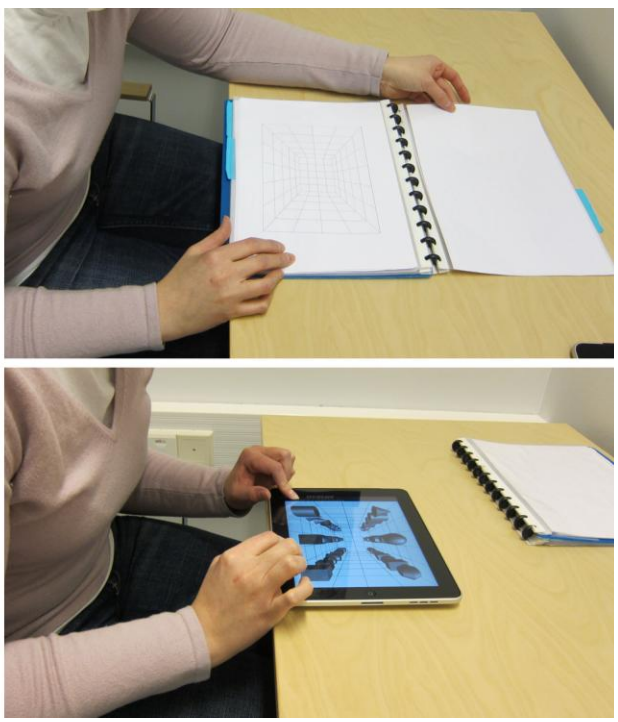


Figure 9. A virtual prototype on a tablet.



A

B

Figure 10. In the validation test, a user made depth level selection tasks A-C with the paper (A) and the virtual tablet prototype (B).

B. Participants' Selections with the Virtual Prototype

In this comparison test, the same depth level selection tasks were conducted with both, paper and tablet prototypes using a total of 34 participants (Fig. 10).

With the virtual prototype, 18% of users selected level 3 from the example without objects (Fig. 11a). When objects were ordered, levels 4 and 5 were both selected by 18 % of participants. 41 % of users selected level 3, when objects were unordered. The majority of users preferred levels from 3 to 5 in all tasks. Deeper levels such as level 30 and infinity were selected by 3-9% of subjects. These selections indicate the same results that were found by the paper prototype. (Fig. 11a.)

In order to compare results with both paper (PP) and virtual prototypes (VP), we combined selections in certain depth levels (Fig. 11b) according to size of hypotenuse in the background grid. The levels were combined as follows:

- level 4 in VP represents the level 3 in PP,
- levels 5-6 in VP represent the level 4 in PP,
- levels 7- 9 in VP represent the level 5 in PP,
- levels 20 and 30 in VP represent level 10 in PP
- levels 10, 11, 14 are not referable to any levels of paper prototype; therefore, they have been moved to NA (not applicable) category. (Fig. 11b)

Results with virtual prototype (Fig. 11b) support our findings gathered by the paper prototype (Fig. 6). Participants in the both tests preferred depth levels 3-5 (Fig. 6 and 11b).

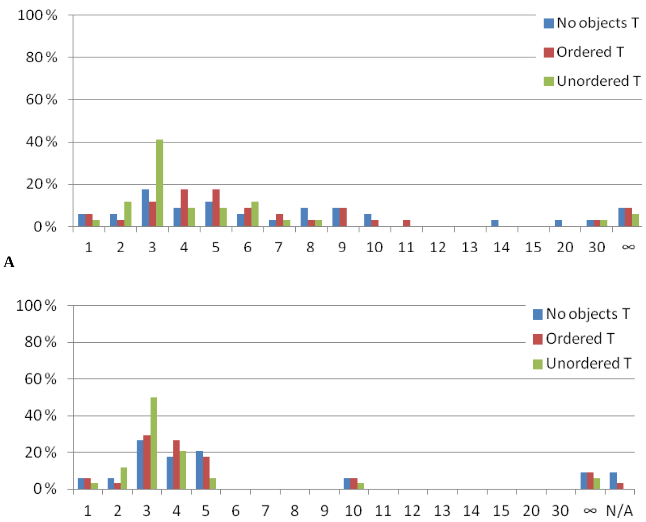


Figure 11. A) Participants' depth level selections with a virtual prototype. B) Virtual prototype's levels combined to paper prototype levels.

V. DISCUSSION

The depth level experiment in the concept evaluation was made in order to get user feedback for the design of 3D UIs in the early design phase. The paper prototype was developed and used for a certain mixed methods evaluation situation. In this context, the paper prototype proved to be a useful method and enabled us to get user feedback for a specific research topic. Especially in mixed methods evaluation procedures, it is important to use different types of prototypes to illustrate design ideas to the users and then communicate about them. This paper contributes UX research by presenting how a fast and low-cost method can be used in the early design phase as a part of mixed methods evaluation procedure.

Also the validation test with the paper and virtual prototypes elicited similar results than we found by the paper prototype in the concept evaluation. In further studies, we will use these findings in new 3D UI designs.

VI. CONCLUSION AND FUTURE WORK

This paper presented how users perceive depth in a 3D UI and which depth level they prefer. Results from the depth level selections with paper and virtual prototypes elicited that users preferred levels from 3 to 5. They liked these depth levels because they thought that they can perceive and select all needed applications easily just from one view without a need for hierarchical menu structures and camera view changes. The level 1 was regarded as boring and too simple. Only a few users preferred the infinity level. This level would be interesting to study more with a running application, because preferences for infinity level can be dependent on the content (e.g., music, photos, contacts). Based on the user evaluations with both paper and virtual prototypes, we propose that the depth levels 3-5 could be the most preferable depth for the 3D UI as a default starting point, depending on the system context. Users could also have a possibility to customize the depth levels when needed.

In further studies, we will use these findings in new designs, and then test how users experience the depth levels with an interactive 3D UI on a touch screen mobile device.

The reason for providing this early phase design and UX evaluation information to 3D UI and HCI research fields is to increase the knowledge of conducting UX studies with time- and cost-effective low-fidelity methods.

ACKNOWLEDGMENT

We thank our funders: Intel, Nokia and Tekes. Thanks to LudoCraft Ldt. for implementing the virtual prototype on the tablet. Warm thanks to our evaluation participants as well.

REFERENCES

[1] A. Agarawala and R. Balakrishnan, "Keepin' It Real: Pushing the Desktop Metaphor with Physics, Piles and the Pen." Proc. CHI 2006, ACM Press (2006), pp. 1283-1292. doi>10.1145/1124772.1124965.

[2] D.A. Bowman, H. Chen, C.A. Wingrave, J. Lucas, A. Ray, N.F. Polys, Q. Li, Y. Haciahmetoglu, J.S. Kim, S. Kim, R. Boehringer, and T. Ni, "New Directions in 3D User Interfaces." The International Journal of Virtual Reality, 2006, vol 5 (2). pp. 3-14.

[3] R. Budiu and N. Nielsen, "Usability of iPad Apps and Websites." www.nngroup.com/reports/mobile/ipad. 30.8.2012.

[4] M. Buchenau and J.F. Suri. "Experience prototyping," Proc. DIS2000, ACM Press (2000). pp. 424-433. doi>10.1145/347642.347802.

[5] Z. Cipiloglu, A. Bulbul, and T. Capin, "A Framework for Enhancing Depth Perception in Computer Graphics." Proc. APGV 2010, ACM Press (2010), pp. 141-148.

[6] A. Cockburn and B. McKenzie, "Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments." Proc. CHI 2002, ACM Press (2002), pp. 203-210.

[7] A. Gotchev, G.B. Akar, T. Capin, D. Strohmeier, and A. Boev, "Three-Dimensional Media for Mobile Devices." Proc. IEEE 99, 4 (2011), pp. 708-741.

[8] ISO 13407: 1999. Human-centred design processes for interactive systems. ISO 13407:1999(E).

[9] ISO DIS 9241-210:2010. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems. International Standardization Organization (ISO). Switzerland.

[10] A. Leal, C.A. Wingrave, and J.J. LaViola, "Initial Explorations into the User Experience of 3D File Browsing." Proc. BCS-HCI'09, ACM Press (2009), pp. 339-344.

[11] J. Light and J.D. Miller, "Miramar: A 3D Workplace." Proc. IPCC 2002, IEEE Press (2002), pp. 271-282.

[12] M. Pakanen, L. Arhippainen, and S. Hickey. "Design and Evaluation of Icons for 3D GUI on Tablets," Proc. MindTrek'12, Oct 2012. ACM Press (2012), pp. 203-206. doi>10.1145/2393132.2393171.

[13] M. Pakanen, L. Arhippainen, and S. Hickey. "Studying Four 3D GUI Metaphors in Virtual Environment in Tablet Context. Visual Design and Early Phase User Experience Evaluation," In Proc. ACHI'13, 2 Feb 24 - Mar 1, 2013 - Nice, France.

[14] D. Patterson, "3D SPACE: Using Depth and Movement for Selection Tasks." Proc. Web3D 2007, ACM Press (2007), pp. 147-155. doi>10.1145/1229390.1229416.

[15] K. Salo, L. Arhippainen, and S. Hickey, "Design Guidelines for Hybrid 2D/3D User Interface on Tablet Devices. A User Experience Evaluation." Proc. ACHI'12, ThinkMind Press (2012), pp. 180-185.

[16] C. Snyder, "Paper Prototyping: The Fast and Easy Way to Design, and Refine User Interfaces (Interactive Technologies)" Elsevier (2003).

[17] A. Vermeeren, E. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila, "User Experience Evaluation Methods: Current State and Development Needs." Proc. NordiCHI 2010. ACM Press (2010), pp. 521-530.

[18] K. Väänänen-Vainio-Mattila and M. Wäljas, "Developing an expert evaluation method for user eXperience of cross-platform web services." Proc. MindTrek'09. ACM Press (2009), pp. 162-169. doi>10.1145/1621841.1621871.

Studying Four 3D GUI Metaphors in Virtual Environment in Tablet Context

Visual Design and Early Phase User Experience Evaluation

Minna Pakanen, Leena Arhippainen and Seamus Hickey
Intel and Nokia Joint Innovation Center/ Center for Internet Excellence
P.O. Box 1001, FI-90014 University of Oulu, Finland
firstname.lastname@cie.fi

Abstract— **In this paper, we focus on a possibility to have a personal 3D GUI inside a virtual environment on tablet device. We describe the visual design process and user experience evaluation of four 3D GUIs in a virtual environment. A user evaluation was conducted by using a structured pair evaluation procedure, where we adapted a concept walkthrough method with non-functional visually high quality prototypes. We found that participants would like to have their personal 3D GUI in a virtual environment. However, the visual design of the 3D GUI should create a secure and private feeling for them. Also, participants did not want the GUI to occlude too much with the background. The visual indication is needed also when a user transfers items from personal GUI to the virtual environment and for showing the user's active position between the GUI and virtual environment. We point out also other issues for interaction and visual designers.**

Keywords: visual design; user experience; 3D GUI; touch screen tablet device, HCI.

I. INTRODUCTION

Three dimensional (3D) collaborative games, for example, Order and Chaos [20], have been developed already for touch screen tablet devices, such as Apple iPad. There have also been interests in bringing virtual environments (VEs) such as Second Life (SL) [18] to the tablets. The first 3D viewer for SL is Lumiya for Android tablets [19]. In 3D VEs, users can see 3D objects and other people's avatars in 3D space. To carry out other activities, such as reading personal emails, browsing files or playing games, a user cannot do this in VE, but she/he needs to switch to another application, which may weaken the 3D environment experience. In this paper, we explore different approaches to that problem by focusing on a possibility to have a personal 3D graphical user interface (GUI) inside a VE. By a personal GUI, we mean a private user interface (UI) showed only to the user, not visible publicly, in contrast to embedded elements in VEs visible to all users.

3D UIs and VEs have been studied over many decades with PCs using several input devices. There is only little research done with tablet devices [23], as earlier studies have focused on larger touch displays such as tables [11][27] and screens on the wall [14]. Bowman et al. [5] define a 3D UI as a UI that involves 3D interaction, which means human-computer interaction (HCI) where a user performs tasks directly in a 3D context. Based on this definition, a 3D interaction can be defined so that it comprises navigation, object manipulation, application control [5][12][29] and visual design [7]. Many earlier 3D UI studies have focused on 3D file browsing, because 3D allows a larger set of items to be displayed at the same time [15][22][8]. Also 3D menus and metaphors have been investigated a lot over the years and the most popular 3D metaphors are: tree, mirror, elevator, book, art gallery, card and the hinged metaphor [10]. As tablet devices have been used for reading books and magazines, a bookshelf metaphor has become quite popular for displaying content, for example, in the Apple iPad [2]. Also 3D carousel metaphors have been under a large interest, both in industry and academy [14][21][30]. Different kinds of 3D and 2½D desktops have been designed and studied as well [1][17][26].

VEs are social in their nature, but if there are personal items in there, then privacy should be clearly visualized to the user. Culnan [9] defines privacy as: *"The ability of individuals to control the terms under which their personal information is acquired and used"*. Privacy is a large research topic, but in this paper, we focus only on visual indication of the privacy in VEs. The prior research has focused mainly on e-commerce applications for selling either real world products or virtual products for avatars [24] or for information exchange between avatars [16]. Butz et al. [6] introduced two visual indication ways (vampire mirror and privacy lamps) for indicating which items are shared and which are private. However, they did not report any tests with users.

The research of personal 3D UI elements (e.g. objects, menu items and files) in a collaborative virtual environment is still lacking from a visual design, user experience (UX) and mobile tablet device point of view. This paper investigates users' expectations of a personal 3D GUI in a collaborative VE in the early design phase and offers preliminary user feedback on visual design and indication of privacy. First, we present the visual design of metaphors and preparation of user experience evaluation study examples. Then, we describe the study with 40 participants conducted with non-functional visually high quality prototypes. Finally, we report the results and point out factors that designers should consider when designing 3D GUIs for VEs on touch screen tablet devices.

II. USER EXPERIENCE CENTERED DESIGN

ISO 9241-110:2010 [13] defines user experience as: "*a person's perceptions and responses that results from the use and/or anticipated use of a product, system or service"*. User experience cannot be designed, because it is in people, but it is possible to 'Design for experiencing' [25].

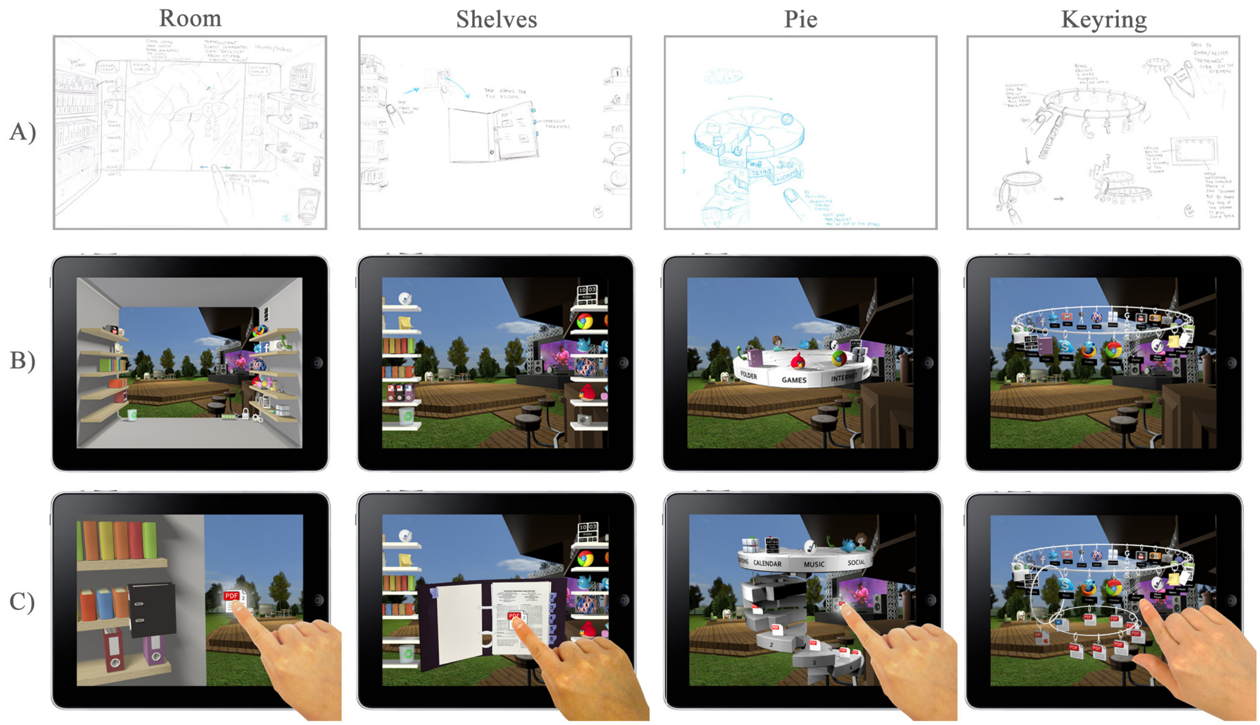


Figure 1. 3D GUI metaphors: A) sketches, B) in a virtual environment and C) with a 'file searching and sharing' use case.

To 'design for experiencing', we used the industrial design process [28]. First, we explored approximately 40 existing 3D UIs and concept designs. Then, based on this benchmarking, literature and lessons learned from our earlier studies [3][23] with 3D UIs we identified three major design goals: *1) Design a 3D GUI in a collaborative virtual environment, 2) Support the use of multiple applications within the 3D virtual environment* and *3) Design for 3D interaction on a touch screen* (for example, object selection from the back rear of a carousel UI). After this, we started the concept design for the 3D GUI metaphors. Finally, we prepared four designs for user experience evaluation.

A. Two Design Phases

We started the first design phase with the preparation of five different styled visual theme boards to help us create visuals for the concepts. These A4 sized boards were composed as a collage of images of different visual forms of the titled style. Then, we had the first brainstorming session utilizing the visual theme boards with two industrial/interaction designers and a UX researcher. This resulted in different kinds of ideas, advice and needs that were written down. Next, we had a one week individual sketching phase when we produced over 100 sketches of 3D GUI metaphors. Then, we had an expert evaluation of the concepts with eight project members.

The second design phase was started with an individual sketching period. We developed the selected concepts of the first design phase further and in more detail. Then, the sketches (approximately 50) were evaluated by eight UI and UX professionals. Based on the evaluation, four 3D GUI metaphors: *Room*, *Shelves*, *Pie*, *Keyring* (Fig. 1, A) were selected for the 3D modeling phase, because they were comparable against each other. *Room* and *Shelves* metaphors have a similar visual style and both of them had a binder metaphor for files but a different amount of icons and depth of space. *Pie* and *Keyring* are examples of the carousel metaphor, but with different visual style, hierarchy level structure and amount of icons.

B. File Searhing and Sharing Use Case

We wanted to have an example with hierarchical structure in the GUIs; therefore, we designed also a step-by-step use case (file searching and sharing*)* for each concept (Fig. 1, C), which are presented in Table 1. The idea of the use case was to search for a PDF file (named as PDF 2), copy and share it to a pre-named contact.

C. Modeling of 3D GUI Metaphors and 3D Icons

The 3D models of the selected concepts were created by using the Blender program. First, we designed and modeled the GUI elements and 3D icons for our 3D GUI metaphors. We selected applications that can be used in the tablet context (e.g. mail, phone, messaging, notebook, radio, maps, contacts, books, browsers, gallery, folder, trashcan, calendar, camera, games, music player and social media services). We had a set of 33 icons to be used in our GUI metaphor designs. The amount of icons in every design varied, because we wanted to have a different evaluation setup for each concept in order to evaluate UI hierarchy structures and the amount of objects displayed in the UI metaphor and on the screen at once. There were 31 icons in the *Room* concept's first view, but in the *Shelves*, there were only seventeen, which are either fully or partially shown icons. The *Keyring*

TABLE I. THE PROCESS OF A FILE SEARCHING AND SHARING USE CASE IN EACH 3D GUI METAPHOR CONCEPT

Steps	The process of a file searching and sharing use case in each 3D GUI metaphors concept			
	Room	*Shelves*	*Pie*	*Keyring*
File searching	**User:** Zooms in with a pinch zoom gesture. **User:** Taps the PDF binder icon. **System:** Opens the binder in the center of the screen. **User:** Taps the 'PDF 2' index marker **System:** Turns the page and the intended PDF is in sight.	**User:** Tap the PDF binder icon on the shelf on the left side of the screen. **System:** Activates and moves a shelf (that the binder is located on) near the center area of the screen and opens the binder in the center of the screen.	**User:** Taps the binder icon which was located on a one piece of the Pie. **System:** The tapped piece of Pie drops one step down and the system opens three sub-pieces of the Pie on the same horizontal level. Three icons are located on top of the pieces; W (Word), PP (PowerPoint) and PDF (2nd hierarchy level). **User:** Taps the PDF icon. **System:** Sub-pieces opens under the Pie GUI in the format of a hierarchical helical stairs (3rd hierarchy level).	**User:** Taps the binder icon. **System:** Vertically orientated sub-ring with three icons; W, PP and PDF appears to hang from the original ring. **User:** Tap the PDF icon. **System:** Another sub-ring opens horizontally to the icon's place **User:** Zooms in (Fig. 1, C, Keyring).
File copying	**User:** Long press the PDF icon **System:** The copied file icon appears on top of the PDF file.	Copying is made similarly as in Room GUI (Fig. 1, C, Shelves).	Copying is made similarly as in the Room GUI.	Copying is made similarly as in the Room GUI.
File sharing by dragging	**User:** Drags the copied file to the other side of the Room (Fig. 1, C, Room) to the contact object (ball), and finally to the chosen contact. **System**: Camera follows the file dragging and zooms in to the contact ball.	**User:** Drags the copied file on another shelf on the other side of the screen with two contact objects (balls) on. **System:** Moves the shelf with contact objects to the center area of the screen. **User:** Drags the copied file to the contact object (ball), and finally to the chosen contact.	**User:** Drag the copied file on a contact piece in the Pie (Fig. 1, C, Pie). **System:** Opens sub-pieces in hierarchical helical stair format, where all the contacts are located on the steps of the 'stairs'. **User:** Drags the copied file to the chosen contact. **System**: Camera follows the file dragging.	**User**: Drags the copied file to the contact object (ball) at the rear of the first hierarchy level ring. **System**: Camera follows the file dragging zooms in the contact object.
Feedback indication to a user	**System:** Shows a tiny version of the icon beside the contact, which disappears when it is sent.	The system indication for sending is done the same way as in the Room GUI.	**System:** Shows a tiny version of the icon beside the contact on the step of the stair, which disappears when the file is sent.	The system indication for sending is done the same way as in Room and Shelves GUIs.

concept included 28 icons and *Pie* ten icons in the first menu hierarchy level. Finally, we made compositions for 'file searching and sharing' use cases by moving and duplicating modeled UI elements.

D. *Preparation of the Prototypes*

We decided to evaluate our four designs as non-functional visually high quality prototypes in as early design phase as possible to get user feedback for the next iteration of our concepts with a fast, easy and cost-effective way.

Because we were interested in finding out the user experiences of the visual aspects, it was important to make high quality looking evaluation examples. Based on our design goals, we wanted to evaluate how users perceive the 3D GUIs in a virtual environment (Fig. 1, B). Therefore, we selected one 3D model of a collaborative looking virtual outdoor music environment from our earlier research work [3] and rendered out one image of it from Blender. Then, we rendered each image of the metaphors with the step-by-step use case and placed them on a VE background in Photoshop. We then added a life-sized 10 inch tablet frame around the images. Finally, we added images of hands which were representing the touch gestures on top of the use case images (Fig. 1, C) and saved the image series as PDFs.

III. USER EXPERIENCE EVALUATION

As we were interested in users' subjective experiences, we conducted the study by using structured pair evaluation and adapted the design walkthrough method in a controlled setting, which lasted from 25 to 59 minutes. We used different methods to gather user feedback and experiences: video recording, semi structured interviewing, observation with writing down comments. First, in the beginning of the evaluation, users filled up a short background questionnaire which had questions about participants' gender, age, prior touch screen and 3D experience. The actual design walkthrough was conducted as follows for each 3D GUI concepts: Showing the 3D GUIs on a 3D VE with the 'file searching and sharing' use case on a life-sized tablet frame as a PDF from a laptop where the moderator changed the image and led the discussion. She asked participants to comment freely what they are thinking and also asked additional questions now and then.

A. *Participants*

In our user evaluation, we had 40 persons of which 63% were male. For recruiting participants, we used an online test user environment and also sent email invitations to friends and colleagues to be distributed. The criterion for selecting participants was that each of them should have at least two months' experience with touch screen devices (mobile phones or tablets). Almost all of the participants (93%) had prior touch screen experience with smart phones and 85% of them had tried or used tablet devices. The participants' age varied from 23 to 52 years, with a mean of 35.

I. FINDINGS

All the material was qualitative, which we analyzed by applying the affinity diagram method [4]. We wrote down participants comments on sticky-notes. Then, we made two analysis rounds for notes and grouped them based on their content. A summary of the analysis is presented in Table 2. In the following subsections, we present participants' perceived aspects and comments on the 3D GUIs in 3D VE.

A. *Perceived Visual Appearance*

The Room metaphor (Fig. 1, B) was considered as a 'homely' GUI where one's own applications are in order. The *Room* metaphor was also called as 'garage' or 'storage', but it was also regarded as childish and funny like 'a toy store'. 18% of the participants thought that the *Shelves* concept was better, clearer, more approachable and pleasurable than the *Room* GUI. The *Pie* GUI metaphor (Fig. 1, B), in its turn, was perceived as interesting, new, exciting and visually attractive. On the other hand, the *Pie* was regarded as an official, masculine and engineering type of object and was called as 'a disk' or 'hard drive'. The visual style of the *Pie*'s plate was perceived to be bulky, chunky and too thick and it was called 'a concrete plate', 'tray', 'puzzle', 'Battle Star Galactic' or 'puck'. It was even suggested that the plate could be translucent. The visual style of the *Keyring* (Fig. 1, B) was considered to be new, different, interesting and fun. On the other hand, one participant commented that it is:*" a moment's wow"*. Compared to the *Pie*, the *Keyring* was regarded as a feminine object and it was called as a kitsch bracelet. It was also referred to movement, for example, to 'a shower curtain rack', 'coat hanger rack', 'mobile', and 'janitor's key ring'. One person even said: *"I don't like if it's swinging"*.

Participants liked the fact that they can easily get an overview of the GUI with one glance, with other GUIs than *the Shelves*. 15% of the participants did not like that all of the icons are not showing. Also, it was perceived as odd and ugly that some of the icons on the shelves were cut in half. In contrast, 30% of the subjects liked the tighter view that the *Pie* concept offered even though there was even less content in sight. *Pie* and *Keyring* were perceived to look like launchers for applications. Participants thought that in the *Room* (18%) and *Keyring* (25%) GUIs, there were too many occluding application icons. It was perceived to be unclear and error prone while making selections.

Participants thought that all GUIs except the *Room*, looked weird and distressing with the virtual environment background, because they seemed to be floating in the air, for example, the *Pie* GUI was perceived as a UFO. Also, one participant commented the meaning of the Pie metaphor because of its location in the 3D environment: *"It looks like a tray when it is located near a bar"*.

B. Perceived 3Dness

When participants evaluated the 3Dness of the concepts, one factor was the depth of the space. Compared to other concepts, in the *Shelves* GUI, there was not enough depth to make it look like a 3D and it was considered to be only a 2D GUI with 3D icons. As one participant commented about it: *"3D icons do not change the UI into 3D"*. Another factor was the perceived interaction. The *Pie* and *Keyring* had the round shape which made them look rotatable; therefore, they were perceived as 3D. Also the icon occlusion was considered to be an important factor for creating a 3D feeling; thus, the *Shelves* concept was not considered to be a 3D GUI. From the users' perspectives, 3Dness is made from occlusion, the shape of the UI and the depth of the space.

C. Perceived Consumption of Space from VE

The occlusion of the virtual environment by the GUI was evaluated by the participants. The *Room* and *Pie* were perceived to consume too much space from the VE. The *Room* GUI was showing the center area, but it was considered more like a little peak view to the VE. With the *Pie* GUI, the situation was quite the opposite; the plate of the *Pie* blocked the center area. In comparison, the Shelves and *Keyring* GUIs were considered to be lighter and airy on VE.

D. Perceived Privacy and Safety

The participants felt more secure with the *Room* concept, because there were walls separating the private area from the public background area. To create a secure feeling, there should be some kind of separation from the background environment. However, with the *Shelves, Pie* and *Keyring* GUIs, participants had concerns for their privacy. For example, one participant commented on the *Pie* GUI: *"If I am in a public virtual space, can other people see my UI?"* The *Shelves* and *Keyring* GUIs were perceived as visually unclear and confusing, because behind the icons and UI elements there were not any visual elements to separate it from the VE. With the *Shelves* GUI, participants wished for a back plate or curtain behind the shelves.

There should also be a clear visual indication for showing the user's active position between the personal 3D GUI and collaborative virtual environment, which could be indicated with color or dimming. Participants thought that a possibility to interact between spaces and share content directly to a friend in the VE was good. On the other hand,

TABLE II. A SUMMARY OF HOW PARTICIPANTS PERCEIVED FOUR 3D UI METAPHORS

UX	Four 3D GUI Metaphors			
	Room	***Shelves***	***Pie***	***Keyring***
Perceived visual appearance	+ homely + things are ordered (garage/storage) + can see all the icons at once - unclear (icons are occluded/ too full) - childish and funny/ toy store	+ clear - shelves are floating in the air (odd) - icons cut in half (ugly) - not possible to see all icons at once - floating in the air (odd)	+ new / exciting / attractive + can see the most important icons at once - bulky / too thick/ chunky - masculine / engineering type / official - floating in the air (odd)	+ new / different / interesting/ fun + can see all the icons at once - full / unclear (icons are overlapping) - feminine/ kitsch bracelet / swinging - floating in the air (odd)
Perceived 3Dness	+ 3D space (Room) + enough depth - icons occluded	- not enough depth = 2D GUI - just 3D icons do not make 3D GUI - no occlusion	+ 3D shape (round) + icons occluded + looks rotatable (interaction)	+ 3D shape (round) + looks rotatable (interaction) - icons occluded
Perceived consumption of space from VE	+ distinct from the background VE - consumes too much space from VE	+ light/airy + does not consume too much space from VE	- consumes too much space from VE	+ light/ airy + does not consume too much space from VE
Perceived privacy and safety	+ clear visual separation from VE (walls) - can other users of VE see the content a shared item	- possible to share something to the VE by accident (no walls) - not clear visual separation from VE - can other users of VE see the content of own GUI or a shared item	- not clear visual separation from VE - can other users of VE see the content of own UI and a shared item	- not clear visual separation from VE - can other users of VE see the content of own UI and a shared item
Perceived ease of use	+ looks simple/ easy to use - require more steps than 2D UI - too long dragging - needs camera & zooming controls	+ no brainer to use + no camera controls required + shorter dragging	+ brainless to use - carousals are difficult - menu hierarchy difficult and messy - too many steps (file search & sharing)	- difficult (can accidently select a wrong icon) - menu hierarchy messy and weird - too many steps (file search & sharing)
Perceived utility by customizaton	+ easy to categorize the content + easy to customize the GUI space	+ easy to categorize the content	+ could work as a launcher	+ could work as a launcher + easy to categorize the content

they were concerned about a possibility to share something into the VE by accident. This could be prevented by giving a user visual indication with a highlight color when something is moved from their personal GUI to the collaborative VE. There were also concerns such as, 'can someone else see a shared file and to whom it is shared to'. The shared content should be invisible to other users and it should look like it is protected for the user who is sharing it and who is receiving it. For example, as one participant suggested: *"Shared file could be protected with a folder?"*

E. Perceived Easy of Use

Even though we did not have a functional prototype, participants commented a lot how they perceived the usability aspects of each GUI metaphors with the 'file searching and sharing' use case. Participants thought that the *Shelves* GUI (Fig. 1, c) was better than other GUIs from the usability point of view. It was perceived to have shorter dragging, simpler hierarchy, fewer steps and camera movements, such as view rotation and zooming in too near to the UI elements. Also, one person said that in the *Shelves* GUI the interaction can be done more *"brainlessly"*.

Even though a tablet is a gestural interface, participants did not like long dragging, because it was perceived difficult and prone to errors, such as an item is dragged to a wrong place. 15% of the participants suggested that instead of long dragging, a copy could be moved to a 'pocket' or virtual USB-memory stick and kept there until sharing. 15% of the participants suggested that the GUI could be intelligent, for example, the target object (in this case a contact), could automatically open beside of the binder while copying.

With the *Pie* and *Keyring* metaphor concepts (Fig. 1, C), the 2nd and 3rd hierarchy levels were found distressing, because there were too many items illustrated at the same time and it looked too messy. When the 3rd hierarchy level opened in the *Keyring* GUI, 30% of the given comments were negative. As one participant commented: *"More and more jingling"*. Also the orientations of the sub-rings was found irritating, against the laws of physics and foolish. Therefore, it was suggested that rings could open horizontally either under the original ring, replacing it or earlier opened rings could move deeper into the space when the new ring opens. With the *Pie* GUI, the 3rd hierarchy level opening as a form of helical stairs was unexpected by the participants and over 50% of the given comments were negative For example, one participant described it: *"It exploded, went broken"*. It was perceived as difficult, hard, complex and distressing. 13% of the participants commented that it looks like endless stairs. The helical stairs structure was perceived to prioritize the content. For example, with contacts, it creates a feeling that some of the contacts are more important than others. With PDF files, the structure was not that irritating, but the amount of items was considered to be critical for the controllability of the GUI. Participants suggested that instead of the 2nd and 3rd hierarchy levels in the *Pie* and *Keyring* GUIs, there could be a similar binder metaphor as in the *Room* and *Shelves.* Other suggestions for the *Pie* included: a drawer opening from it or another *Pie* could open under the first one. With *the Keyring*, there could be a binder metaphor or file cabinet instead of the 2nd and 3rd hierarchy levels.

F. Perceived Utility by Customization

Users thought that customization would be interesting and useful within all of the GUIS. The simplest thing with the UI customization is to let the users to adjust the amount of icons. Also, some people liked to categorize the GUI content. For example, with the *Keyring*, participants wanted to categorize icons in groups or pile them in stacks. With the *Room* and *Shelves* GUIs, participants would have liked to organize their icons by placing work and leisure items on different sides of the room or on different shelves. Also 30% of the participants were interested in decorating the *Room* GUI space, for example, with wallpapers.

II. DISCUSSION

Our study indicated that the participants would like to use their personal 3D GUI in a collaborative virtual environment, but they must know what others can see from their private or shared 3D UI components. Therefore, the visual design of the GUI is important for creating a secure and private look for the 3D GUI. It can be designed with visual elements, such as walls or curtains, which will distinguish the GUI and its elements from the background environment. However, these elements should not excessively occlude the virtual environment; therefore, they could be also translucent.

There should also be a clear visual indication for showing the user's active position between the personal 3D GUI and collaborative virtual environment, which could be indicated with color or dimming. Also, when something is transferred from the personal GUI to the public virtual environment, there should be a visual indication for the user, for example, a highlight color. Virtual environment also affects how participants perceive the visual design of the 3D GUI. Participants thought that all GUIs except the *Room* looked weird and distressing with the VE background, because they seem to be floating in the air and were unclear looking.

From the users' perspectives, 3Dness in 3D GUI is made from icon occlusion, the shape of the UI and the depth of the UI space. The shape of the GUI and amount of the icons depends on the user's personal preferences. One user wishes to see all of the applications at one glance and another one would just want to see only the most important applications in their GUI. Therefore, there should be different kinds of GUI designs available to the participants. Also the hierarchy structure does not have to continue similarly through the hierarchy levels; it is more preferable is to use flat hierarchy on touch screen devices. Users need to have a possibility to organize icons, UI elements and decorate the GUI space as they wish. Therefore, the customization of the GUI is important for the users.

Even thought this evaluation had limitations from the interaction point of view, it nevertheless provided useful information to us for the next iteration of the relevant concepts. It is not possible to evaluate touch screen interaction with a non-operational prototype, but it is possible to show the designed interaction ways and discuss about them with participants. Therefore, we can now give these early design phase user perceptions to designers and developers to make more pleasurable VEs in the future.

III. CONCLUSION AND FUTURE WORK

In this paper, we focused on a possibility to have a personal 3D GUI inside a VE on tablet device. Therefore, we did the visual design process of 3D GUIs with 'file searching and sharing' use cases. Then, we had an early design phase UX evaluation of four 3D GUIs in a virtual environment background with 40 participants. This evaluation with non-functional visually high quality prototypes gave us a lot of user feedback for the design. Participants liked a possibility to have their personal 3D GUI in a virtual environment. However, the visual design of the 3D GUI should create a secure and private feeling for them. The secure feeling can be designed with visual elements, such as walls or curtains, which will distinguish the GUI and its elements from the background VE. However, the GUI should not occlude too much with the background VE. There should be visual indication presented when a user transfers items from the personal GUI to the virtual environment and also for showing the user's active position between the GUI and VE. We pointed out also other design issues relating to the visual design of the personal GUI for interaction and visual designers of 3D GUIs and virtual environments.

In the future studies, we need to design and implement a functional prototype to test interactions as well. Especially we need more information about the dragging action length and camera movements.

IV. ACKNOWLEDGEMENTS

This work was carried out in Tekes Chiru project. Thanks to Julianna Hemmoranta for 3D modeling and to our evaluation participants.

REFERENCES

[1] A. Agarawala, and R. Balakrishnan, Keepin' it real: pushing the desktop metaphor with physics, Piles and the Pen. Proc. CHI 2006, ACM Press (2006), 1283-1292.

[2] Apple iPad Bookshelf. http://tinyurl.com/95hdvcz 14.12.2012.

[3] L. Arhippainen, M. Pakanen and S. Hickey. Designing 3D Virtual Music Club Spaces by Utilizing Mixed UX Methods: From Sketches to Self-Expression Method. Proc. MindTrek 2012, ACM Press (2012), 178-184.

[4] H. Beyer, and K. Holtzblatt. Contextual design: defining customer-centered systems. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, 1998.

[5] D.A. Bowman, J. Chen, C.A. Wingrave, J. Lucas, A. Ray, N.F. Polys, Q. Li, Y. Haciahmetoglu, J. Kim, S. Kim, R. Boehringer and T. Ni. New Directions in 3D User Interfaces. The International Journal of Virtual Reality, 5, 2 (2006), 3-14.

[6] A. Butz, C. Beshers and S. Feiner. Of Vampire mirrors and privacy lamps: privacy management in multi-user augmented environments. Proc. UIST '98, ACM Press (1998), 171-172.

[7] Z. Cipiloglu, A. Bulbul, and T. Capin. A Framework for Enhancing Depth Perception in Computer Graphics. Proc. APGV 2010, ACM Press (2010), 141-148.

[8] A. Cockburn, and B. McKenzie. 3D or not 3D? Evaluating the effect of the third dimension in a document management system. Proc. CHI 2001, ACM Press (2001), 434-441.

[9] M. J. Culnan. Protecting privacy online: Is self regulation working? Journal of Public Policy and Marketing 19, 1, (2000), 20-26.

[10] A. Gotchev, G.B. Akar, T. Capin, D. Strohmeier, and A. Boev. Three-Dimensional Media for Mobile Devices. IEEE 99, 4 (2011), 708-741.

[11] M. Hancock, T. ten Cate and S. Carpendale. Sticky Tools: Full 6DOF Force-Based Interaction for Multi-Touch Tables. Proc. ITS '09, ACM Press (2009), 133-140.

[12] C.A. Hand. Survey of 3D Interaction Techniques. Computer Graphics Forum, 16, 5 (1997), 269-281.

[13] ISO 9241-210:2010. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems. International Standardization Organization. Switzerland.

[14] G. Jacucci, A. Morrison, G. Richard, J. Kleimola, P. Peltonen, L. Parisi, and T. Laitinen. Worlds of Information: Designing for Engagement at a Public Multi-touch Display. Proc. CHI 2010, ACM Press (2010), 2267-2276.

[15] A. Leal, C.A. Wingrave, and J.J. LaViola. Initial Explorations into the User Experience of 3D File Browsing. Proc. HCI 2009, ACM Press (2009), 339-344.

[16] C.Y. Lee and M. Warren. Security issues within virtual worlds such as Second life. Proc. 5th Australian Information Security Management Conference, Research online (2007), 142-151.

[17] J. Light, and J.D. Miller. Miramar: A 3D Workplace. Proc. IPCC 2002, IEEE Press (2002), 271-282.

[18] Linden Lab. Second Life (2012). http://secondlife.com/

[19] Lumiya viewer for Second Life. http://www.lumiyaviewer.com/. 14.12.2012

[20] Order and Chaos (2012). http://orderchaosonline.com/. 14.12.2012.

[21] D. Patterson. 3D SPACE: Using Depth and Movement for Selection Tasks. Proc. Web3D 2007, ACM Press (2007), 147-155.

[22] G. Robertson, J. Mackinlay, and S.K. Card. Cone Trees: Animated 3D Visualizations of Hierarchical Information. Proc. CHI '91, ACM Press (1991), 189-194.

[23] K. Salo, L. Arhippainen and S. Hickey. Design Guidelines for Hybrid 2D/3D User Interfaces on Tablet Devices - A User Experience Evaluation. Proc. ACHI 2012, ThinkMind Press (2012), 180-185.

[24] R.-A. Shang, Y.-C. Chen, and S.-C. Huang. A private versus a public space: Anonymity and buying decorative symbolic goods for avatars in virtual world. Computers in Human Behavior, 28, 6 (2012), 2227–2235.

[25] E.B-N. Sanders. Virtuosos of the Experience Domain. http://www.maketools.com/papers-2.html. 14.12.2012.

[26] L. Staples. Representation in Virtual Space: Visual Convention in the Graphical User Interface. Proc. INTERCHI '93, ACM Press (1993), 348-354.

[27] O. Ståhl, A. Wallberg, J. Söderberg, J. Humble, L. Fahlén, A. Bullock, and J. Lundberg. Information Exploration Using the Bond. Proc. CVE '02, ACM Press (2002), 72-79.

[28] K. Ulrich. and S. Eppinger. Product Design and Development. McGraw-Hill, New York, USA, 2008.

[29] A. van Dam, A.S. Forsberg, D.H. Laidlaw, J.J. Jr. LaViola. and R.M. Simpson. Immersive VR for scientific visualization: a progress report. Computer Graphics and Applications, 20, 6 (2000), 26-52.

[30] S. Wang, M. Poturalski, and D. Vronay. Designing a Generalized 3D Carousel View. Proc. CHI 2005, ACM Press (2005), 2017-2020.

Towards a 3D User Interface in a Touch Screen Device Context

An Iterative Design and Evaluation Process

Leena Arhippainen, Minna Pakanen, Seamus Hickey
Intel and Nokia Joint Innovation Center/ Center for Internet Excellence
P.O. Box 1001, FI-90014 University of Oulu, Finland
leena.arhippainen@cie.fi, minna.pakanen@cie.fi, seamus.hickey@cie.fi

***Abstract*— This paper presents a design and evaluation process in the early design phase towards a three dimensional user interface on a touch screen mobile device with service multitasking use cases. Our Service Fusion concept is an outcome of this iterative concept design and evaluation process. In this paper, we present briefly four 3D user interface concepts, evaluated with users and experts. Two of these concepts are also implemented on a touch screen mobile device. We also present user experience findings and how they have supported our early phase design process. The Service Fusion demonstration of 3D multitasking shows multiple services running in a 3D city model, which allows users to purchase movie tickets, listen to music and find local bars and stores by drag and drop gestural controls. Evaluation findings indicate that people are interested in service multitasking and dragging and dropping interaction model in 3D user interfaces in the tablet device context. However, the 3D space context will affect how important and useful users perceive services to them.**

Keywords-3D UI; concept design; user experience; touch screen mobile device; tablet.

I. INTRODUCTION

Research in three dimensional (3D) user interfaces (UIs) has a long history ranging from various areas such as Augmented Reality (AR), Cave Automatic Virtual Environment (CAVE), Virtual Environments (VEs), 3D Authoring tools, simulation and entertainment [4][12][15][20]. Even though 3D UIs have been studied for many decades, they are still actively researched [3][7][14][18]. There has been a resurgence of interest in 3D technology in general, especially among stereoscopic 3D movies and commercially available devices with autostereoscopic displays (e.g., LG Optimus 3D phone and Nintendo 3DS game console). Similarly, the introduction of commercially available 3D input sensors over the last few years (e.g., the Wii remote, EyeToy and MS Kinect) have put 3D UI technology into everyday use. Furthermore, 'touching the 3rd dimension' technologies have seen a host of new research based on the above devices [13].

3D is a hot topic nowadays. Even though there is a lot of hype around the 3D topic, there still exists a big question: 'What benefit 3Dness can provide?' Our hypothesis is that a 3D UI could be beneficial provided that it can enhance the user's tasks; therefore, we have investigated 3D UIs in tablet device contexts from the multitasking point of view. In our case, multitasking means that in the same virtual 3D environment, a user can use several services or applications at the same time by dragging and dropping gestural controls. Currently, in tablet devices, users need to switch between applications in such a way that they can view only one application at a time.

The rest of the paper is organized as follows. Section II describes an iterative design and evaluation process towards a 3D UI on a touch screen mobile device. Also the user experiences of our 3D UI concepts and prototypes are briefly presented. Then, the findings are summarized and discussed in Section III, whereas Section IV concludes the paper and provides ideas for future work.

II. EARLY PHASE CONCEPT DESIGN AND EVALUATIONS

ISO 9241-110:2010 standard defines user experience (UX) as: *a person's perceptions and responses that results from the use and/or anticipated use of a product, system or service* [11]. According to the UX professionals, users' experiences should be evaluated before, during and after the use [21] with different types of design examples. In the human-computer interaction (HCI) research field, there is a need for systematic (low-cost) methods and guidelines for UX evaluation [21], especially in the early phase of the development process [22]. In iterative and agile development processes, time- and cost-effective UX studies give important UX feedback for the design [22]. In our 3D UI UX research, we have utilized the early phase design and evaluation approach and conducted various studies, which are briefly described in the in this paper. Finally, we present our Service Fusion concept which we implemented and tested with users.

A. 3D Desktop UI Concept with a File Sharing Use Case

We designed a 3D Desktop UI concept with an abstract virtual environment, which included similar activities that one can have with mobile phones (calls, clock, calendar). We created an example concept of a 'Call handling' scenario, where a user is sharing a file during the conference call as shown in Fig. 1. The virtual participants of the conference call were represented as avatars and the normal call handling functionality occurred by interacting with the avatars. Additionally, as 3D is often cited as means to provide a better visualization of information to the user, additional applications, blogs, web links and certain files are organized and viewable.

After the concept was created, we downloaded sets of example images (20 images in total) into a tablet device and presented the scenario to the participants as a low-fidelity virtual prototype. Participants hold the tablet on their hands

ISBN: 978-1-61208-250-9

and explained how they would use the UI (e.g., tap or drag an icon) (Fig. 2). We evaluated this concept with four users. The concept evaluation gave information about how users perceive 3D UIs and what benefits such UIs could provide to them. Also issues of avatars (what they are for?), privacy (can everyone see my objects in the 3D space?) and multitasking possibilities were raised. Users were interested in the idea that the system could show several services, for example calendar and news feeds, simultaneously with the conference call and users could easily find and allocate meeting times and share files. User experiences from this small scale user study are reported in [5] and 3D visual indications aspects are presented in [16].

Figure 1. An example image of the 3D Desktop UI concept: a person is sharing a PDF file during a conference call.

Figure 2. A user is explaining how he would use the 3D UI.

B. *3D Office UI Concept with Chat & Calendar Use Case*

In the early phase of the project, we designed various 3D UI concepts with different virtual environments, e.g., music and office (meeting room) environments. Fig. 3 illustrates one example of our 3D UI, where a user is communicating in the collaborative 3D Office environment, and services (e.g., a clock and calendar) are shared with other meeting members. In this example, users were able to drag calendar events to that chat for proposing events and meetings (Fig. 3). The other person was then able to accept the event by a touch gesture (e.g., rotate the event block in the chat text by swiping it up and down on the y-axis).

This 3D Office UI concept was studied as a part of a larger concept evaluation procedure [2][17] with a total of 40 subjects (Fig. 4). Findings indicated that the visual design of

Figure 3. An example image of the collaborative virtual 3D Office UI concept with Chat and Calendar use case: dragging a calendar event to the chat discussion.

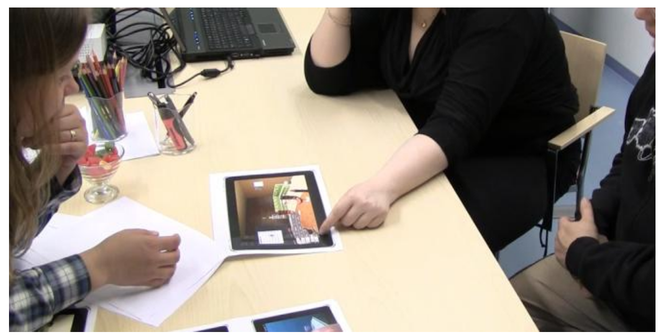


Figure 4. A moderator (left) introduces to the participants the 3D Office UI concept with visually high quality images printed on a paper. A user is commenting the concept (right) in a pair evaluation session.

the VE has a strong impact on how users perceive the whole 3D UI. Changing the 3D virtual environment changes also the users' expectations about a concept or a system, for instance, what kind of behaviour and content they expect from the system. The visual appearance of the 3D virtual environment should be designed for supporting the use context, for example, home, work, leisure, game or map [1]. The use context shown in the 3D VE influences on how important users perceive the service for them, for instance, participants wondered how this 3D Office UI support their work activities (e.g., booking meetings, updating calendar, share files, manage contacts). Therefore, for the next concept, we selected a 3D map as a space context and designed different service multitasking use cases.

C. *A Context-Aware Map Based 3D UI Prototype*

Based on our previous concept studies, we found out that the visual appearance of the 3D virtual environment can set the context of use for the user. In this study, we wanted to exploit this fact as part of our context-aware map based 3D UI prototype implementation. While there are many potential environments to use, a geographical based virtual environment was chosen as the space context. One rationale for this is that geographical based map services are well known, there are many services which have a map aspect and there are a large number of services which include GPS data, which should make it easy to integrate new context- aware services.

This prototype was implemented on Android OS platform on a tablet device. The scenario of the prototype presents a use case whereby a user wants to understand the geographical context of a story in order to get a better understanding of the issues involved. The start view of the UI is shown in Fig. 5A.

A

B

Figure 5. A) 3D UI screen with a map with a number of embedded 3D objects. B) A map of Fukushima with local news.

The UI shows a geographical space, news feeds, calendar functions, contact cards, weather and time. Scenario includes the following steps:

1. The user reads a short news feed about the Fukushima power plant accident. Important data in the news feed is represented as objects.
2. Not knowing where Fukushima is, or what the impact on the local surrounding area is, the user selects the 'Fukushima word' from the news feed by long pressing it and then dragging it to the main map area.
3. When the Fukushima text is dropped on the map environment, the camera transition to the Fukushima virtual environments starts (Fig. 5B).
4. The local weather at Fukushima is shown.
5. Additional news events, pulled from Google news service, are automatically displayed.

This prototype implementation showed the general idea of using a geographical context as a virtual environment. One benefit in this 3D UI example for a user is that he/she can use a map as a search engine easily by dragging and dropping the interesting news items. With tablet devices, typing can be slow and error prone; therefore, we can assume that 'drag and drop' action could speed the use of 3D UI. However, as the prototype implementation was very limited, we tested it only by an expert evaluation (Fig. 6). To prepare the prototype for actual user testing, we developed a new version of the 3D Map UI with and several services.

Figure 6. In a contex-aware map based 3D UI prototype, a user is able to select words (e.g., Fukushima) from news feed and then drag and drop it onto the map environment for getting more information about that news.

D. 3D Map UI with Service Fusion Prototype

After our first Fukushima scenario, we decided to create another use cases relating to context-aware services that citizens could use with their tablets. One idea of our Service Fusion concept is that a user can use several services in the same 3D environment view without a need to switch between application windows. Our aim is to ease the user's tasks by automating some actions with service communications and decreasing a user's typing effort. Therefore, we implemented the following features into the Service Fusion prototype with the 3D city model (Fig. 7):

- **Cinema service:** A user can view the daily timetable of cinema and drag the movie from the list to the personal shopping cart and book and pay for the movie ticket by dragging their credit card object to the payment field (the real payment function not implemented) (Fig. 8).
- **City event Live:** Users can view live events from the local city live calendar.
- **Local News Feeds**: Users can view local news feeds on the embedded board objects.
- **Music service:** Users can view the current song list of a bar and then drag the song name to a radio icon. When the song is dropped on the radio icon, it launches a search for the song from Grooveshark [8]. Users can also drag a song from the list and drop it onto the map to find music stores in the city (a map as a search engine).

This concept was implemented on the realXtend [19] Tundra platform running on Linux Ubuntu on 12 inch touch screen laptop computer. The 3D city model is a high fidelity model and currently covers nine city blocks. The technical implementation is illustrated in the [6][10].

This Service Fusion prototype was tested with ten users. In the evaluation, we were particularly interested in users' subjective experiences of the 3D UI and services. We studied the following topics:

- how users perceive and understand different types of icons. For instance, personal overlaid icons (calendar, shopping cart, radio) are fixed to the camera position and they have a different visual style than in 16 icons (e.g., bars, cinema, theatre, restaurants), which are embedded on the top of the 3D city model and are publicly shown to all users. Also two news tables were embedded in the 3D city map.
- how users experience the dragging and dropping controls with a touch screen mobile device. These results are presented in more detail in [9].
- how users experience the service fusion idea, where a user is able to use several services (e.g., cinema purchase, music listening, music search) in the same 3D virtual environment. In current tablet devices, users have to switch between different applications (e.g., YouTube, Web browser, calendar). These results are presented in more detail in [9].

Figure 7. A Service Fusion 3D UI in a 3D city map: personal calendar, e-commerce services and live information from local bars, cinema and newspapers.

After a short familiarization with a visual appearance of the 3D UI, the users were asked to perform the following tasks:

1. Buy a movie ticket for today:
   a. Find cinema information (cinema icon and movie)
   b. Book movie (drag movie to a shopping cart)
   c. Fill ID info (drag an ID card to a form)
   d. Select a seat (tap the seat)
   e. Pay the ticket (drag a visa card to a form). (Fig. 8)
2. Listen to a music:
   a. Find the bar: Otto K. (icon on the top of the map)
   b. Look at its information board (playing list).
   c. Find out how to listen to it (drag on a radio icon).
3. Search music stores:
   a. Select the song from the playing list (info board)
   b. Find out how to locate a music store which sells CDs and merchandises from that band. (Fig. 9-10).

User tests with this prototype gave us a lot of important feedback about icons design, virtual environment design and

Figure 8. A user is dragging a Visa card onto the payment area.

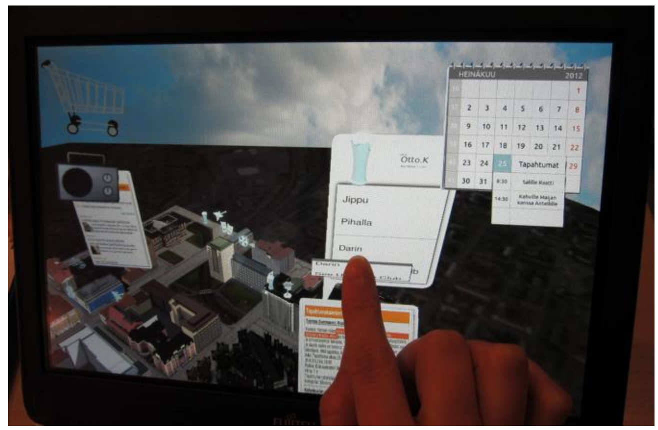


Figure 9. A user presses a song item on the bar's playing list, when the song text (band's and song's name) pop-up as a 3D box, then the users holds the song box and and starts to drag it onto the 3D map for searching a music store (User task 3).

Figure 10. When the song is dropped on the map, the song box disappear and the 'Music store' icon on the top of the map growns and its information board opens, which shows the song information (e.g., price).

service multitasking activities. Findings from this study elicited that the users understood the difference between overlaid icons and icons on top of the 3D city. Especially icons on the top of the 3D city were perceived to belong to the city, a.k.a. are services or stores of the city in real life. Some users understood that an overlaid calendar icon is a public event calendar of the local city. This perception can be a consequence of the test device and setup, because some users may think that the device could not be his/her own device and own personal calendar.

Fig. 11 presents the Likert Scale answers about the icon design (1='totally disagree', 3='cannot say', 5='totally agree'). The participants understood a meaning of the icons (4.3) on the top of the 3D city. They did not think that there are too many icons in the scene (1.3). When a 3D virtual environment includes interactive objects and icons, it is important to design them so that they distinguish enough from the background (e.g., 3D city model) and a user can understand how he/she can interact with them. According to the questionnaire, users thought that icons distinct enough from the background space (3.8). However, the average 3.8 shows that icons could be decided even more distinguishable. Also results relating to appearance of interactive icons (3.4) indicate that embedded icons should be designed so that a user can understand how he/she can interact with them. However, users thought that the visual appearance of icons was consistent (3.9).

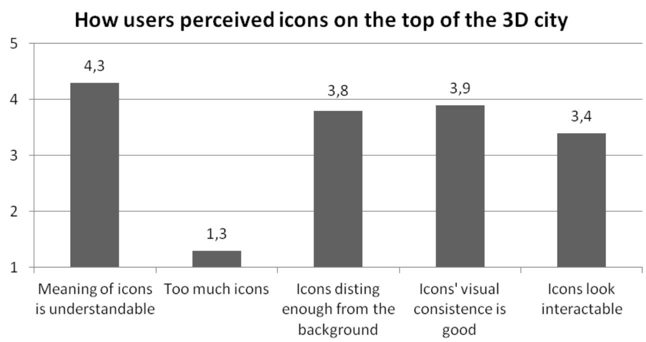


Figure 11. Users' perceptions on icons on the top of the 3D city (embedded icons) gathered by a Likert Scale questionnaire.

The users' experiences on the cinema service were very positive. They liked the idea that they can easily drag and drop items on the 3D UI. The users regarded buying a movie ticket by dragging it on the shopping cart as interesting, easy, fast, useful and natural [9]. Even though the users liked the idea of buying tickets by dragging and dropping ID and credit cards, security issues should be designed so that users can feel the payment system as reliable and secure.

In the music service task, a dragging a song item on the radio icon was natural to users. Instead, dragging the same icon on the 3D city map was not natural at all [9]. Users did not find this feature by themselves. They found it only by accident or after the moderator's tip. However, after users knew this action, half of them liked the idea of using a map as a search engine. The use of a 3D virtual environment, in our case the 3D city map, as a search engine or other interactive platform is a very important topic to study more from the user experience point of view.

III. Summary and Discussion

In this paper, we have shown how we have used different concept designs in an iterative design and evaluation process. In all of these examples, we have investigated how a 3D UI could be useful for users, a.k.a. what additional value the 3D UI could provide to users. Our hypothesis was that service multitasking in one 3D scene (virtual environment) could be useful for users. In the 3D Desktop UI concept, we studied how users can use a 3D environment for sharing files in a conference call setting. In this study, the users liked the idea that they can easily share files by dragging and dropping them onto the friend's avatar or some marked area. They were also willing to share social media and calendar information with persons in the conference call. Also in this example, we noticed that because the visual style of the virtual environment was very abstract, users did not associate the context of use very strongly to certain situations; instead, they perceived the concept example relating to the communication situation.

Next, we designed different virtual environment contexts, e.g., music and office, and studied those examples with users. In the evaluations, users strongly tied examples relating to the certain usage situations, e.g., in leisure or professional activities. The virtual background of the 3D UI had a strong influence on how users perceived the whole concept and its purpose. In this study we also investigated how the users perceive the use of shared services in a collaborative situation.

Based on the findings from these both studies, we started to design how we could utilize a geographical background as a part of the 3D UI and how users would perceive service multitasking in this context. The user study elicited that a geographical based 3D UI with context-aware service multitasking can have benefits for users by making the use of services easier and faster in a 3D virtual environment in the tablet device context. Users liked the idea of dragging and dropping objects to the services. Also users’ perceptions on icons elicited visual design directions for interactive icons embedded in the 3D environment. We will study these topics more with functional prototypes with different 3D environments, with different sets of services.

This iterative design and evaluation process has provided us valuable user feedback that we will use for future designs. One interesting future topic is how we can utilize parts of a 3D environment as interactive platforms (e.g., search engines, shared and public areas). In addition, in the future studies, we will implement and test a prototype, where a user can switch between different 3D virtual environments through teleporting points (e.g., doors or holes). In this study, we will investigate the user experience of the different 3D virtual environments and transitions between them. Also, we will study the visual indication of 3D icons.

IV. Conclusion and Future Work

In this paper, we presented our iterative design and evaluation process for the studying the user experience of different 3D UI examples from the service multitasking point of view. As an outcome of this process, we present our Service Fusion concept and prototype implementations. In UX evaluations, we have investigated how the users perceive different 3D UI design examples and what benefits they could provide to the users. In our 3D UI concepts, the value for users would be the multitasking approach which supports a simultaneous and parallel use of different applications in one screen view in the touch screen mobile device context. In the future, we will conduct more user studies for 3D UIs with different sets of virtual environments and services. The reason for providing this early phase design and UX evaluation information to 3D UI and HCI research fields is to increase the knowledge of conducting UX studies before, during and after the functional prototype.

ACKNOWLEDGMENT

We thank our funders: Intel, Nokia and Tekes. We would also like to thank Julianna Hemmoranta, Jukka Kaartinen, Erno Kuusela, Antti Karhu and Ludocraft Ldt. for their assistance. We want to express our gratitude to Grooveshark and Kutalab Ltd. for providing us with access to their services. Warm thanks to all test users as well.

REFERENCES

[1] L. Arhippainen, M. Pakanen, and S. Hickey. "Preliminary Design Guidelines for Tablet's 3D GUIs Based on the Five UX Studies," Proc. Designing Interactive Systems (DIS 2012), Poster no. 136., New Castle, UK, June 23-15.2012.

[2] L. Arhippainen, M. Pakanen, and S. Hickey. "Studying User Experiences on Depth in 3D User Interface by a Paper Prototype as a Part of the Mixed Methods Evaluation Procedure," Proc. ACHI' 13, 2 February 24 - March 1, 2013 - Nice, France.

[3] D.A. Bowman, J. Chen, C.A. Wingrave, J. Lucas, A. Ray, N.F. Polys, Q. Li, Y. Haciahmetoglu, J. Kim, S. Kim, R. Boehringer, and T. Ni. "New Directions in 3D User Interfaces," The International Journal of Virtual Reality, 2006, vol 5 (2), pp. 3-14.

[4] F.P. Brooks. "What's Real About Virtual Reality?" IEEE Computer Graphics and Applications, 1999, vol 19, 6, pp. 16-27.

[5] Chiru blog: Studying UX of Call handling with 3D UI in tablet. http://chirublog.org/2011/05/20/studying-ux-of-call-handling-with-3d-ui-in-tablet/. 20.5.2011.

[6] Chiru blog: Service Fusion at MUM 2012. http://chirublog.org/2012/12/20/service-fusion-at-mum-2012/. 20.12.2012.

[7] A. Gotchev, G.B. Akar, T. Capin, D. Strohmeier, and A. Boev, Three-Dimensional Media for Mobile Devices. IEEE 99, 4 (2011), pp. 708-741.

[8] Grooveshark: http://grooveshark.com/. 30.10.2012.

[9] S. Hickey, M. Pakanen, and L. Arhippainen. "A 3D UI for Service Multitasking in a 3D City Map," Proc. OZCHI 2012, 26-30 November 2012, Melbourne, Australia, pp. 208-211.

[10] S. Hickey, M. Pakanen, L. Arhippainen, E. Kuusela, and A. Karhu. "Service Fusion: An Interactive 3D User Interface," Proc. MUM 2012. Article No. 53. 4-6. December 2012. Ulm, Germany. doi>10.1145/2406367.2406432

[11] ISO DIS 9241-210:2010. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems. International Standardization Organization (ISO). Switzerland.

[12] G. Kurtenbach, G. Fitzmaurice, T. Baudel, and B. Buxton. "The design of a GUI paradigm based on tablets, two-hands, and transparency," Proc. CHI'97, Atlanta, Georgia, USA, March 22-27, 1997. ACM New York, NY, USA, pp. 35-42.

[13] J. LaViola, and D. Keefe, "3D spatial interaction: applications for art, design, and science," In ACM SIGGRAPH 2011 Courses, Article No. 1. Vancouver, British Columbia, Canada, August 7-11th. New York, NY, USA. ACM Press (2011). doi>10.1145/2037636.2037637

[14] A. Leal, C.A. Wingrave, and J.J. LaViola. "Initial Explorations into the User Experience of 3D File Browsing," Proc. BCS-HCI'09, ACM Press (2009), pp. 339-344.

[15] J. Light, and J.D. Miller. "Miramar: a 3D workplace," Proc. IPCC 2002, IEEE Press (2002), pp. 271-282.

[16] M. Pakanen, L. Arhippainen, and S. Hickey. "UX Based Design Recommendations for Interactive 3D Objects in 3D UI on Tablet Devices," In: Anderson, K., Arhippainen, L., Benko, H., de la Rivière, J-B., Häkkilä, J., Krüger, A., Keefe, D., Pakanen, M., and Steinicke, F. Proc. ACM CHI Workshop on The 3rd Dimension of CHI: Touching and Designing 3D User Interfaces (3DCHI), Austin, Texas, USA. May 5-10, 2012. pp. 91-94. http://www.cie.fi/chi-workshop.html

[17] M. Pakanen, L. Arhippainen, and S. Hickey. "Studying Four 3D GUI Metaphors in Virtual Environment in Tablet Context. Visual Design and Early Phase User Experience Evaluation," Proc. ACHI' 13, 2 February 24 - March 1, 2013 - Nice, France.

[18] D. Patterson. "3D SPACE: Using Depth and Movement for Selection Tasks," Proc. Web3D 2007, ACM Press (2007), pp. 147-155. doi>10.1145/1229390.1229416.

[19] realXtend: http://realxtend.org. 30.8.2012.

[20] S. Reifinger, F. Laquai, and G. Rigoll. "Translation and rotation of virtual objects in Augmented Reality: A comparison of interaction devices," In IEEE SMC, 2008, pp. 2448-2453.

[21] A. Vermeeren, E. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila. "User Experience Evaluation Methods: Current State and Development Needs," Proc. NordiCHI 2010. ACM Press (2010), pp. 521-530.

[22] K. Väänänen-Vainio-Mattila, and M. Wäljas. "Developing an expert evaluation method for user eXperience of cross-platform web services," Proc. MindTrek 2009. ACM Press (2009), pp. 162-169. doi>10.1145/1621841.1621871.

Subjective Usability of Speech, Touch and Gesture in a Heterogeneous Multi-Display Environment

Arnoud P.J. de Jong Susanne Tak Alexander Toet Sven Schultz Jan-Pieter Wijbenga Jan van Erp
TNO, The Netherlands
{arnoud.dejong, susanne.tak, lex.toet, sven.schultz,jan_pieter.wijbenga, jan.vanerp}@tno.nl

Abstract— **Several interaction techniques have been proposed to enable transfer of information between different displays in heterogeneous multi-display environments. However, it is not clear whether subjective user preference for these different techniques depends on the nature of the displays between which information is transferred. We explore subjective usability of speech, touch and gesture for moving information between various displays in a heterogeneous multi-display environment, consisting of a multi-touch table, a wall-mounted display and a smartphone. We find that subjective user evaluation of the various interaction techniques depends on the combination of displays being used. This implies that the type of display combination should be taken into consideration when designing interaction techniques for the transfer of items between displays in a heterogeneous multi-display environment. Also, gesture based interactions were judged more acceptable when they involved holding a mobile phone, probably since this provided a cue explaining the action.**

Keywords- large display; multi-display environment; multi-touch table; smartphone; speech; gestures

I. INTRODUCTION

Distributed computing environments (e.g. meeting rooms, collaborative work spaces) are increasingly populated with many heterogeneous display devices like smartphones and tablets (providing small personalized displays), tabletop displays (facilitating collaboration between small groups), and large size displays (for information presentation to larger groups).

A frequent task in these heterogeneous multi-display environments is moving objects between displays [1]. A typical example is the exchange of files (e.g. images or documents), either from a mobile device to a tabletop (for group activities), from a tabletop or mobile device to a large wall display (for public presentation), or from a large screen or tabletop to a mobile device (for personal use).

To minimize user error and workflow interruptions the techniques for cross-display interaction should be simple and intuitive, requiring minimal physical and cognitive effort. Also, they should be "socially acceptable" (i.e. they should not make the user feel uncomfortable).

Although several cross-display interaction techniques have been proposed, it is still unknown if subjective user preference for these different techniques depends on the nature of the selected display pairs. Therefore, the current study assesses subjective user preferences for gesture, touch and speech based interaction techniques in a setting that integrates a multi-touch table with a wall-mounted display and a smartphone. To the best of our knowledge this is the first study to investigate these three interaction modalities in the same heterogeneous multi-device collaborative computing environment. In the rest of this paper, we will first discuss related work on interaction and navigation techniques used in multi-display environments. Then we will describe a user study that we performed to assess the subjective usability of gesture, touch and speech techniques for this purpose, and we will present the results of this study. Finally, we will present our conclusions and we will provide suggestions for future research.

II. RELATED WORK

Multiple or distributed display environments present new challenges for interaction and navigation. Currently the interaction with heterogeneous multi-display environments is still dominated by a single-user single-display paradigm: the user interacts with one display at a time, using an interaction technique that is considered most appropriate for a particular type of display. Available interaction techniques are typically keyboard, touch, gesture, or speech based. Keyboard input has long been the standard but can be too slow and cumbersome in dynamic environments. Touch based interaction has become a popular interaction technique for devices like mobile phones, tablets and interactive tabletops. Although it is generally fast, it is only suitable for direct interaction at close range. Gestural interaction has gained popularity since the application in interactive computer games (Wii, Microsoft's Kinect system). This technique is more appropriate for direct interaction with large displays that can be operated from a distance [2]. Speech based interaction has become a common direct interaction technique for in-car navigation devices and hands-free phone systems, and might gain in popularity with the increasing availability of voice operated smartphone apps (e.g. via Siri on the iPhone). Similar to touch, speech interaction is only suitable for direct interaction at close range.

Recently several new interaction techniques have been proposed to move objects between tabletop displays and mobile phones [3-6], tablets [1, 7] or hand held devices in general [8], between hand-held and large displays [9, 10], or between any of these devices [11]. Many of these techniques are gesture based.

Speech and gestures complement each other and (when used together) can create an interaction technique that is more powerful than either modality alone. Speech interaction is suited for descriptive techniques, while gestural interaction

 ISBN: 978-1-61208-250-9

is ideal for direct manipulation of objects [12]. Speech allows interaction with objects regardless of their degree of visual exposure (occlusion). It appears that users prefer using combined speech and gestural interaction over either modality alone when handling graphics manipulation [13].

III. USER STUDY

To assess user preference for different interaction techniques in a heterogeneous multi-display environment, we performed a study in which users transferred items (photographs) between different types of displays, using gesture, touch and speech techniques. Subjective user experience was quantified through semantic questionnaires.

A. *Method*

A setup with a multi-touch table, a wall-mounted screen and a smartphone was used (see Figure 1). Participants performed the same task (i.e. transferring a photograph from one display to another) several times for four different display pairs, using various techniques.

B. *Task and interaction techniques*

Participants were requested to send a photograph from one display to another using various techniques. The four display pairs were (touch)table to screen, table to mobile (phone), screen to mobile, and mobile to screen. The interaction techniques were as follows:

Table to screen
Speech (A1): Select photo and say "*send to screen*".
Touch (A2): Drag photo to a window entitled '*Screen'*.
Gesture (A3): Select photo and point at the screen.

Table to mobile
Speech (B1): Select photo and say "*send to Harry*".
Touch (B2): Drag photo to a window entitled '*Harry'*.
Tangible (B3): Place mobile on table and drag photo to it.

Screen to mobile
Speech (C1): Start voice command by saying "*screen*", then say "*send to Harry*".
Gesture (C2): Hold phone as if taking a photo of the screen.

Mobile to screen
Speech (D1):Start voice command by dragging finger downwards over the screen and say "*send to screen*".
Touch (D2): Press send button below photo and select the '*Screen'* menu item.
Gesture (D3): Point phone at the screen.

C. *Design and procedure*

Participants completed tasks in all display pair × technique combinations, using a repeated measure within-subject design. The four different display pairs were presented in random order. For each display pair, the experimenter first demonstrated the various interaction techniques. To familiarize the participants with the interaction techniques they were given the opportunity to practice. Next, participants were requested to perform the interaction (transferring the photo from one display to the other) five consecutive times for each interaction technique. After completing the tasks for a particular interaction technique a questionnaire was administered verbally. The questionnaire contained seven statements, each related to a particular aspect of perceived usability:

1. I could execute the task without thinking (*without thinking*);
2. The interaction was intuitive (*intuitive*);
3. The interaction felt unnatural (*unnatural*);
4. The interaction was tiring (*tiring*);
5. The system responded quickly (*responsive*);
6. The interaction is complex (*complex*);
7. The interaction is error-prone (*error-prone*).

Participants rated their agreement on a 5-point Likert scale ranging from "completely disagree" to "complete agree". In addition, for each display pair, participants were asked to rank the three (two for the screen to mobile display pair) techniques from most to least preferred. On average, participants completed the experiment in 60 minutes.

D. *Participants*

Twenty-one people participated in the user study (12 male, 9 female, 20-57 years old, average age 27). Fourteen participants owned or regularly used a device that uses touch input (e.g., smartphone or tablet). The experiment was undertaken with the consent of each participant. Participants were paid 30 euro for their participation.

IV. RESULTS

The questionnaire results show significantly different ratings for all questions (Friedman test, $p< .001$ for *without thinking*, *intuitive*, *unnatural*, *complex* and *error-prone*, $p< .01$ for *tiring* and *responsive*). Post hoc pairwise comparisons (Wilcoxon with Bonferroni correction, $\alpha=.05$) were used to examine the questionnaire results in more detail. Where relevant, these results are discussed in the sections on each of the display pairs below.

The seven questions were converted to one overall subjective usability score by converting all answers to scores ranging from -2 to 2: completely agreeing on a positively framed statement such as without thinking was scored 2, while the reverse coding scheme was applied to negatively framed statement such as *unnatural*. Next, these scores were summed and averaged for all conditions (mean Cronbach's alpha .76), leading to overall scores in the range [-2,2]. The results are shown in Figure 2.

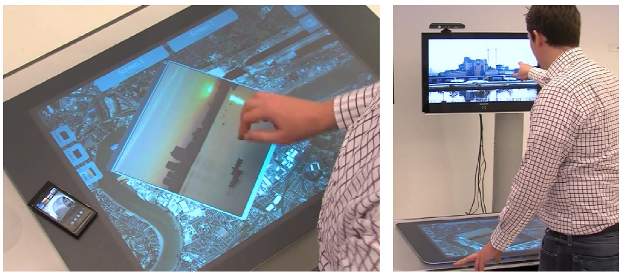


Figure 1. Experimental setup; interacting with the multi-touch table (left) and the screen (right)

 ISBN: 978-1-61208-250-9

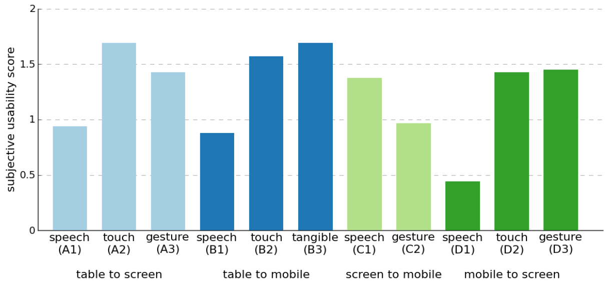


Figure 2. Subjective usability scores for the 11 experimental conditions.

The subjective usability scores (measured across the 11 different conditions) were analyzed using a one-way repeated-measures ANOVA [14]. The results show a main effect for condition (F10,200=11.1, p<.001), which was further analyzed using post-hoc pair-wise comparisons with Bonferroni correction (α=.05). These results are discussed in the sections on each of the display pairs below.

A. *Table to Screen*

Analysis of the subjective usability scores revealed significant differences between the speech technique (A1) and the touch technique (A2) (p<.05). Participants had a clear preference for the touch technique; 17 participants preferred touch the most, 4 participants preferred gesture the most, and no participants preferred speech the most (Figure 3, X2(2)=22.57, p<.001). Further analysis of the qualitative remarks by the participants revealed that 5 participants explicitly reported that they thought speaking the commands out loud was awkward. In general, people were very positive about the touch technique. Finally, the gesture technique was also well received by the participants, with 4 participants reporting that the method was fun to use. However, 4 other participants remarked that they considered this particular gesture (stretching their arm and pointing) to be embarrassing or too “commandeering”.

B. *Table to mobile*

Analysis of the subjective usability scores revealed significant differences between the speech technique (B1) and the touch technique (B2) (p<.05), and the speech technique (B1) and the tangible technique (B3) (p<.01). This is also reflected in the preferences for the various techniques; 15 participants preferred the tangible technique the most, 5 participants preferred gesture the most and 1 participant preferred speech the most (X2(2)=14.86, p<.001). Furthermore, 18 participants preferred speech the least (Figure 3). In particular, the results showed that people considered the speech technique for this task type significantly more unnatural than both the touch and the tangible techniques (both p<.05), which was also evident from the qualitative remarks made by the participants (similar to the table to screen task, 4 participants reported feeling awkward when speaking the commands out loud). The tangible technique was very well received, with 10 participants calling it “fun” or “cool”.

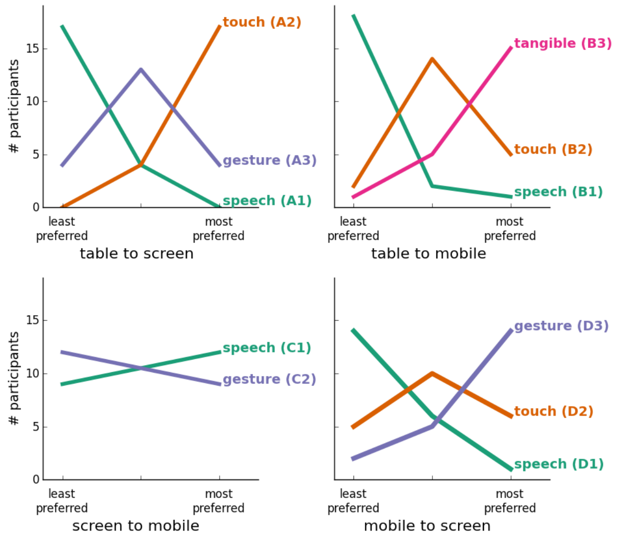


Figure 3. Number of participants preferring a certain interaction mode (touch, gesture, tangible, speech) for each of the 4 different device pairs.

C. *Screen to mobile*

Analysis of the subjective usability scores revealed no significant difference between the speech technique (C1) and the gesture technique (C2). There was also no significant preference for either of the techniques; 12 participants preferred speech the most, and 9 participants preferred gesture the most (X2(1)=0.2, p>.6). This is in contrast to the other display pairs, where speech was generally the least preferred technique (Figure 3). Analysis of the qualitative remarks made by the participants reveals that participants had no reservations about “talking to the screen”. This is in contrast to the table to screen/mobile tasks, where people felt uncomfortable using speech.

D. *Mobile to screen*

Analysis of the subjective usability scores revealed significant differences between the speech (D1) and touch technique (D2) (p<.05), and between the speech (D1) and gesture technique (D3) (p<.01). This is reflected in the user preferences: 14 participants preferred the gesture technique the most, 6 participants preferred touch the most, and 1 participant preferred speech the most (X2(2)=12.29, p<.01).

V. CONCLUSIONS AND FUTURE WORK

Overall, the results show that subjective user preference for the interaction techniques depends on the type of task.

In general, the speech technique was not very well liked: people often reported feeling embarrassed when speaking commands out loud. Note that the voice commands were in English, which probably introduced an extra degree of difficulty for the participants, who were not native English speakers. In addition, some participants had strong preconceptions about speech being an inappropriate interaction mode. Speech was only the preferred interaction mode when the large screen was the target of the speech command (C1; see Figure 2). Possibly, speech was deemed acceptable in this case because there are fewer viable interaction alternatives (the screen is too distant to touch).

Though gesture-based interaction techniques were generally well received (except for some participants having issues with the required physical effort), there are some interesting differences between the different tasks. In particular, the use of gestures was not preferred in the table to screen task (A3), while it was positively received on the mobile to screen task (D3, see Figure 3). This is interesting because the gestures are similar for both task types, with the key difference being whether the participant is holding an object (smartphone) or not. Participants regularly felt uncomfortable or embarrassed using a gesture for the table to screen task, but not for the mobile to screen task. Possibly, holding an object that provides a clear visual cue explaining the user's actions makes gesture-based interaction more acceptable [15, 16]. Gesture-based interaction techniques may become more acceptable when people are allowed to use tangible objects, perhaps even if these objects serve no technical purpose (i.e. a dummy object). We note that the subjective usability scores showed no difference between the various gesture conditions, but the user preference rankings did (see Figures 2 and 3). Also, qualitative remarks made by the participants suggest a difference between the various gesture conditions. This suggests that our questionnaire was incomplete in that sense, and that future user evaluations should explicitly address embarrassment. More specifically, in future studies we intend to test whether participants are primarily spatial, verbal or object oriented.

This study focused on the subjective evaluation of different interaction techniques. In future studies we also intend to register objective performance measures (e.g. the time it takes to perform the different actions).

Finally, we note that we only investigated single users interacting with single display pair, with only the experimenter present. Future research could investigate settings with multiple users and/or more complex display combinations.

REFERENCES

[1] Nacenta, M.A., Aliakseyeu, D., Subramanian, S. & Gutwin, C. (2005). A comparison of techniques for multi-display reaching.Proc. CHI '05 (pp. 371-380). New York, USA: ACM.

[2] Bragdon, A. & Ko, H.-S. (2011). Gesture Select: acquiring remote targets on large displays without pointing.Proc. CHI '11 (pp. 187-196). New York, USA: ACM Press.

[3] Chehimi, F. & Rukzio, E. (2010). Throw your photos: an intuitive approach for sharing between mobile phones and interactive tables.Proc. Ubicomp 2010 New York, USA: ACM Press.

[4] Shirazi, A.S., Döring, T., Parvahan, P., Ahrens, B. & Schmidt, A. (2009). Poker Surface: combining a multi-touch table and mobile phones in interactive card games.Proc. MobileHCI '09 New York, USA: ACM P.

[5] Schmidt, D., Chemini, F., Rukzio, E. & Gellersen, H. (2010). PhoneTouch: a technique for direct phone interaction on surfaces.Proc. ITS '10 New York, NY, USA: ACM.

[6] Yoo, J.-W., Hwang, W., Seok, H., Park, S.K., Kim, C., Choi, W.H. & Park, K. (2010). Cocktail: exploiting bartenders' gestures for mobile interaction. Int.J.Mob.Hum.Comp.Interaction, 2(3), 44-57.

[7] Bader, T., Heck, A. & Beyerer, J. (2010). Lift-and-drop: Crossing boundaries in a multi-display environment by Airlift.Proc. AVI'10 (pp. 139-146). New York, USA: ACM Press.

[8] Olsen, D.R., Clement, J. & Pace, A. (2007). Spilling: expanding hand held interaction to touch table displays.Proc. TABLETOP '07 (pp. 163-170). Los Alamitos, CA, USA: IEEE Press.

[9] Hutama, W., Song, P., Fu, C.-W. & Goh, W.B. (2011). Distinguishing multiple smart-phone interactions on a multi-touch wall display using tilt correlation.Proc.CHI '11 (pp. 3315-3318). New York, USA: ACM Press.

[10] Jeon, S., Hwang, J., Kim, G.J. & Billinghurst, M. (2006). Interaction techniques in large display environments using hand-held devices.Proc. VRST '06 New York, NY, USA: ACM Press.

[11] Kray, C., Nesbitt, D., Dawson, J. & Rohs, M. (2010). User-defined gestures for connecting mobile phones, public displays, and tabletops.Proc. MobileHCI '10 (pp. 239-248). New York, NY, USA: ACM Press.

[12] Oviatt, S. (1999). Ten myths of multimodal interaction. Comm.ACM, 42(9), 74-81.

[13] Hauptmann, A.G. & McAvinney, P. (1993). Gestures with speech for graphic manipulation. Int.J.Man-Machine Studies, 38(2), 231-249.

[14] [14] Carifio, J. & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. Medical Education, 42(12), 1150-1152.

[15] Rico, J. & Brewster, S. (2010). Usable gestures for mobile interfaces: Evaluating social acceptability.Proc. CHI '10 New York, NY: ACM Press.

[16] Montero, C.S., Alexander, J., Marshall, M.T. & Subramanian, S. (2010). Would you do that?: Understanding social acceptance of gestural interfaces.Proc. MobileHCI '10 New York, USA: ACM Press.

 ISBN: 978-1-61208-250-9

Dynamic Gesture Recognition Based on Fuzzy Neural Network Classifier

Ching-Han Chen[1] Nai-Yuan Liu[3]
Department of Computer Science and Information Engineering, National Central University
No.300, Jhongda Rd., Jhongli City, Taiwan
[1 3]pierre@csie.ncu.edu.tw

Kirk Chang[2] Gimmy Su[4]
Delta Electronics, Inc.
No 3, Dongyuan Road, Jhongli City, Taiwan
[2 4]kirk.chang@delta.com.tw

Abstract— **This paper presents a dynamic gesture recognition method based on the combination of the fuzzy features of the dynamic gesture track changes and the fuzzy neural network inference system. This method first classified the dynamic gestures roughly into circular gestures and linear gestures. Further, gestures were classified narrowly into up, down, left, right, clockwise, and counter-clockwise gestures. These six dynamic gestures, which are commonly used in IP-TV controlling, were introduced as the recognition goal in our dynamic gesture recognition system. The results show that this method has a good recognition performance and fault tolerance, and more applicable to real gesture-controlled human-computer interactive environment.**

Keywords—**- gesture recognition, fuzzy system, neural network**

I. Introduction

Human-computer interaction is a study discussing the interaction between users and computer systems. Through human-computer interaction, computer systems and users are able to communicate with each other. Human-computer interaction makes the communication easier, fits in with users' needs, and enhances the interaction [1]. In recent years, the methods of human-computer interaction constantly improved. "The more intuitive and natural technologies will replace mouses and keyboards," Bill Gates mentioned on BBC News websites [2], which means touching, visual, and voice interfaces will be more and more important. This is so-called the natural user interface. Thus, gesture recognition has become very popular as a human-computer interaction in recent years.

Human hands, through each joint, are able to make various combinations of actions. A wide range of human-computer interactions [1] are developed by gestures. Gestural human-computer interaction interface can basically be divided into static gestures [2,3] and dynamic gestures [4]. Static gesture is a static image with static gesture information. Dynamic gesture is composed of a series of continuous static gesture images. The information includes the changing of the gestures. This paper focuses on the dynamic gesture recognition. Dynamic gesture recognition analyses the locus of hands or the changing information of the hand movements in continuous images. Presently, the main methods of dynamic gesture recognition are Hidden Markov Model (HMM) [5], Dynamic Time Warping [2], and Neural Network [6].

Track-based dynamic gesture [7] is often unable to meet the diversity of the real world if it only applies pre-established gestures database for feature extraction, training and classification. For example, it may cause identify failure by the actual gesture waving in different backgrounds from the gestures in database. Besides, the same gesture made by different people would be different, either. Even the same gesture made by one person at different times would show significant differences.

We propose a fuzzy neural network classifier taking dynamic gesture locus as feature. First, we extract the data of the gesture locus. Then we classify gestures into linear gestures and circular gestures by two fuzzy neural network classifiers. Finally, we precisely classify the six dynamic gestures, which are commonly used in IP-TV controlling, by the changing of the tracks.

II. Feature Extraction of Gesture Locus

First, we take a gesture locus as a virtual ellipse and define three features in continuous hand locus block coordinates, which are the ratios of major and minor axes, difference between major and minor axes, and the frequency difference between clockwise and counter-clockwise.

A. Major axis and minor axis ratio

In a sequence of M continuous images, we record the center coordinates of the hand area in each image. Then, in these M data, we find out the maximum value of x (X_{max}), the maximum value of y (Y_{max}), the minimum value of x (X_{min}), and the minimum value of y (Y_{min}). Furthermore, we define the major axis as a and the minor axis as b of the hand area, as shown in Fig. 1 (a) and make $\frac{a}{b}$ as t_1.

$$\begin{cases} a = X_{max} - X_{min} \\ b = Y_{max} - Y_{min} \end{cases} \quad (1)$$

B. Major axis and minor axis difference

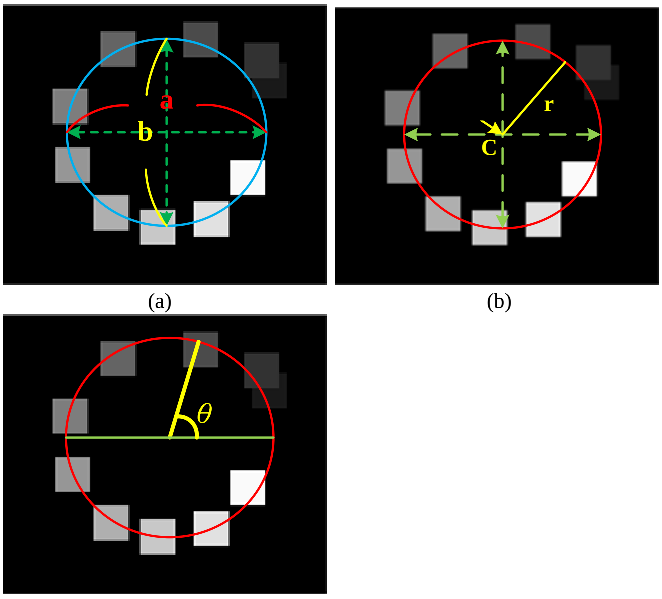
We define t_2 as the difference between major axis and minor axis (a-b).

C. Clockwise and counter-clockwise frequency

After Xmax, Ymax, Xmin, Ymin are found, we calculate the intersects between a and b as the center C of the virtual circle. And take $r = \frac{a+b}{2}$ as the radius of the virtual circle, shown in Fig. 1 (b). Then we calculate the angle θ between the horizontal line across C and the line from C to each center of the hand area, as shown in Fig. 1 (c).

We can decide the gesture is clockwise or counter-clockwise by M different angels (θ). And record the numbers of times that clockwise (f_{CW}) and counter-clockwise (f_{CCW}) occur. We make the clockwise and counter-clockwise frequency t_3 as $f_{CW} - f_{CCW}$.

However, the three characteristics of the continuous gesture locus, t_1, t_2, and t_3, are not well-defined, susceptible to interference. Thus, we fuzzify the three characteristics and put them into the Fuzzy Neural Network to classify.



(a) (b)

(c)

Fig. 1 Definition of the three characteristic parameters of the gesture locus (a)major and minor ratio (b)major and minor difference (c)clockwise/Counter-clockwise frequency

III. Fuzzy Neural Network Classifier

Fuzzy Neural Network (FNN) combines the concept of fuzzy theory and neural network. In recent years, its applications are proposed continuously [8,9]. In general, the neural network and fuzzy theory both are able to estimate the system without mathematical models. They are used to simulate the function of the human brain. The neural network imitates the message passing mechanism in brain cells and the fuzzy system simulates the mental status and psychological reasoning of human.

Basically, in order to achieve the purpose of learning, a function estimation of the neural network is basic on the training data input, output values and the connecting bonding parameters between neurons which are adjusted through repeated error corrections. It is worth noting that there is no way to know the intranet architecture. We cannot directly encode the normal if-then rules in the network. The only method is giving a large number of training data to the system. Compare with the artificial neural networks, the fuzzy systems can directly encode the expert knowledge values in the system, with high tolerance. Thus, through the complementary nature of the neural network and fuzzy system, this combination system has the advantages of both. [10].

A. Fuzzy System

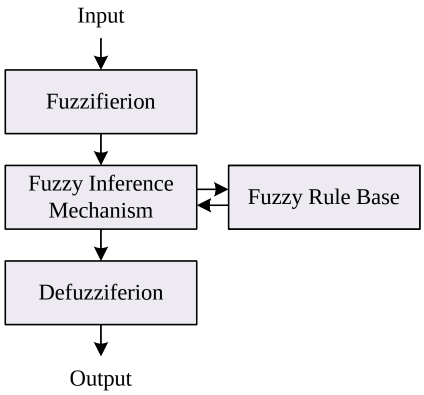


Fig. 2 The basic architecture of the Fuzzy System

Fuzzifierion is to translate the precise input data into a set of syntax fuzzy information. Fuzzy sets are composed of a number of membership functions. In this study, we use the trigonometric functions as membership functions.

The fuzzy inference mechanism is the core of the fuzzy system. It simulates the decision model of human thinking by processing the approximate reasoning or fuzzy inference. The classifier has two inputs, and the fuzzy rules expressed as follows:

R^j:

$$\begin{aligned}&\text{If } x_1 \text{ is } A_1^j \text{ and } x_2 \text{ is } A_2^j \\ &\text{then } X \text{ belong class } y_j \text{ with } CF = CF_j \end{aligned} \quad (2)$$

R^j represents the j-th fuzzy rule. A_1^j and A_2^j represent the fuzzy sets where j = 1 to n with n rules. $y_j \in y\{1,2,\ldots,M\}$ is the output of the j-th fuzzy rule, which is one for the M classes, CF_j represents the reliability of the fuzzy rule R^j. We apply the Generalized Modus Ponens and Max-Min composition operation, and the fuzzy inference outputs expressed as follows:

$$\mu_y(y) = \max_j \{\mu_{y,j}\} \mid y_j = y \quad (3)$$

where,

$$\mu_{y,j} = \mu m_{A,j}(x) \cdot CF_j \quad (4)$$

$$\mu m_{A,j}(x) = \min\left\{\mu_{A_1^j}(x_1), \mu_{A_2^j}(x_2)\right\} \quad (5)$$

The process of converting the fuzzified value into a specific value called defuzzifierion. Formally, the outputs from the fuzzy inference may be the fuzzy sets or specific values. If the result after inference is a fuzzy set, the median method and the center area method are applied to obtain specific outputs. In this study, a single neuron is used to defuzzify, which means to connect the output value u_j of each fuzzy rule directly to the neuron and output the defuzzified result by the neuron.

B. The Single Neuron

In Artificial Neural Network (ANN), neurons can accept the input signal from linked unit and calculate the input value with the bond value. Then it decide whether to pass the signal down to the neuron in the next layer by checking if the calculated value is greater than a threshold. Back Propagation Network (BPN) can make the output values and the training data expectations as feedback to the network. So the network is able to adjust the bond values by the excitation from

 ISBN: 978-1-61208-250-9

external environment. This kind of learning ability can be used to generalize the fuzzy rules in the fuzzy neural network.

Considering the demand for memory and computation time in a real-time system, we decide to use single neuron. Although it is impossible to have all the ability to map nonlinear function, this can be resolved after the operation of the front-end fuzzy system. The single neuron accepts *p* external inputs, calculated with *p* bond values, and input the result to the transfer function $y = \varphi(v)$. Then it outputs the classification inference probability. The formula is as follows:

$$v = \sum_{i=0}^{p} w_i u_i \qquad (6)$$

$$y = \varphi(v) = \frac{\alpha_2}{1 + e^{-\alpha_1 v}} \qquad (7)$$

Where α_1 controls the shape of the function, α_2 adjusts the size of scale. Fig. 3 shows the combination of the fuzzy system and the single neuron.

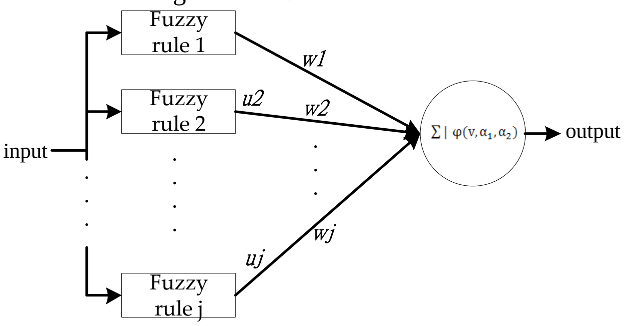


Fig. 3 The artificial neural network classifier combined with single neuron

IV. DYNAMIC GESTURE RECOGNITION

We use the six dynamic gestures, which are considered as the most common using gestures in IP-TV controlling interface, as the recognition goals of the classifier,

- Gesture waving to the left
- Gesture waving to the right
- Gesture waving up
- Gesture waving down
- Gesture waving clockwise
- Gesture waving counter-clockwise

Based on the fuzzy neural network classifier (Fig. 3), we establish two fuzzy neural network classifiers, the circular gesture classifier and the linear gestures classifier. The input features of the linear gesture classifier are the ration and the difference of the major axis and the minor axis. The input features of the circular gesture classifier are the clockwise/counter-clockwise frequency and the ratio of the major and minor axes. In fuzzy neural network, every inference output from the fuzzy rules is linked to a neuron.

It fuzzifies the input feature values by the fuzzification model. Then process the fuzzy inference by the fuzzy rules. Finally, output the classified results computed by the neuron. Fig. 4 shows the two fuzzy neural classifier architectures. O_1 is the inference probability of the linear gestures; O_2 is the inference probability of the circular gesture.

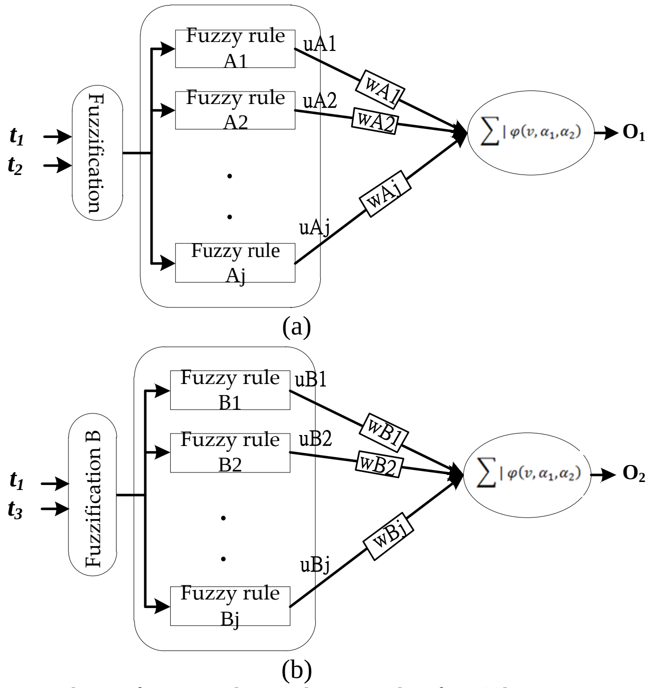


Fig. 4 The two fuzzy neural network gesture classifiers (a)linear gesture classifier (b)circular gesture classifier

The fuzzy neural network classifier can roughly determine the dynamic gestures to be either linear gestures or circular gesture. Further, in accordance with the positional relationship of the gesture trajectory coordinates, it determines the dynamic gesture to be up, down, left, right, clockwise, or counter-clockwise.

It determines the relative position in x-direction and y-direction of two adjacent hand areas of the *M* hand area coordinates, which means to make the determination by the *j*-th point coordinates (x_i, y_i) and the (*j*+1)-th point coordinates (x_{i+1}, y_{i+1}), where $y = 0,1,\ldots,M-1$. The thing need to be noticed is that the waving gestures taken by the camera is left-right reversal to the direction of the user's real waving direction. We take the user’s direction as determination and the formula is as following:

$$\begin{cases} C_{left} = C_{left} + 1 & ,if\ x_{i+1} > x_i \\ C_{right} = C_{right} + 1 & ,if\ x_i > x_{i+1} \\ C_{up} = C_{up} + 1 & ,if\ y_i > y_{i+1} \\ C_{down} = C_{down} + 1 & ,if\ y_{i+1} > y_i \end{cases} \qquad (8)$$

$$C_{sum} = C_{left} + C_{right} + C_{up} + C_{down} \qquad (9)$$

Where, C_{left}, C_{right}, C_{up}, C_{down}, are the times of waving in different directions.

Combining with the times of waving clockwise $f_{CW}(C_{CW})$ and counter-clockwise $f_{CCW}(C_{CCW})$, we define the probabilities of the six dynamic gestures and the formula is as following:

$$\begin{cases} p_1 = \frac{C_{left}}{C_{sum}} \times O_1 \\ p_2 = \frac{C_{right}}{C_{sum}} \times O_1 \\ p_3 = \frac{C_{up}}{C_{sum}} \times O_1 \\ p_4 = \frac{C_{down}}{C_{sum}} \times O_1 \\ p_5 = \frac{C_{CW}}{C_{CW} + C_{CCW}} \times O_2 \\ p_6 = \frac{C_{CCW}}{C_{CW} + C_{CCW}} \times O_2 \end{cases} \quad (10)$$

Where, $p_{1,...,}p_6$ are the probabilities of different dynamic gestures. O_1 and O_2 are the outputs of the linear classifier and the circular classifier. The flow chart of the process recognizing is as following:

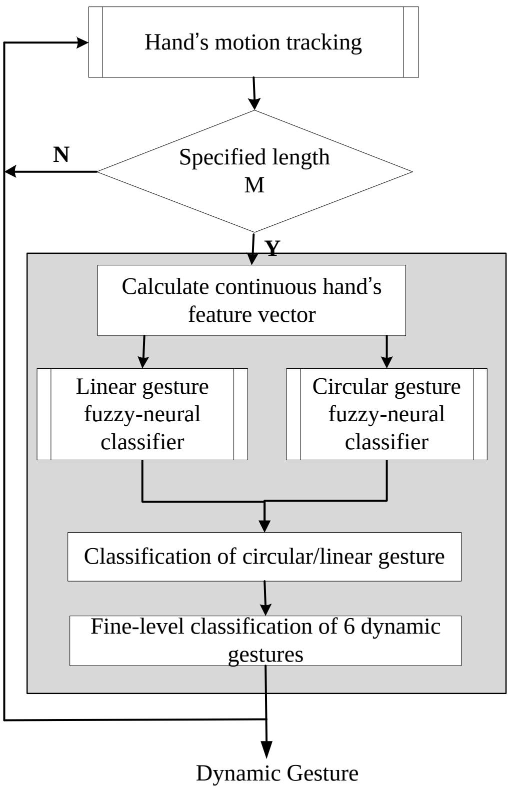


Fig. 5 The flow chart shows the process of recognizing the dynamic gestures by the fuzzy neural network

V. EXPERIMENT

We create a set of continuous images of dynamic gestures in a database. 10 different people were recorded under the six dynamic gestures, each gesture 10 times, a total of 600 gestures and each gesture has 100 samples. The image sequence length to recognize the dynamic gestures is 10 frames.

TABLE I
THE RESULTS OF DYNAMIC GESTURE RECOGNITION

Gesture	**Recognizable Probability**
Left	92%
Right	95%
Up	89%
Down	79%
Clockwise	96%
C-Clockwise	95%
Average	**91%**

The experimental results show that the up and down gesture recognizable probability is lower. Thus, we analyze the gestures to be mistaken as what kinds of gesture. The experimental results are shown in Table 2.

TABLE III
ERRONEOUS GESTURE RECOGNITION ANALYSIS

Gesture	**Recognized as**		
Up	Clockwise	Counter-Clockwise	Down
	3%	2%	6%
Down	Clockwise	Counter-Clockwise	Up
	6%	6%	9%

We observed that the experimenter's waving habits are related to the erroneous recognition. For example, when some people wave upward, their hands fall naturally after they finish the waving action. However, the up-down waving is recorded and recognized. So does the down gesture. The down-up gesture is recorded, too. And the erroneous recognition occurs.

Shan et al. [11] applied a MHI conversion to the tracking results by Mean Shift Embedded Particle Filter (MSEPF), which has the advantages of both Particle Filtering and Mean Shift. Then, it extracts features by using the seven constant torque variables that Hu proposed. Further, the template matching method proposed by Bobick & Davis [12] was applied and Mahalanbis Distance was taken as matching basis. We experiment with the same database and compare the identification performance between this study and Shan algorithm. The results are shown in Table 3. The recognition rate is higher than the method Shan et al. [11] proposed, although the average recognition time of this algorithm is longer.

TABLE IIIII
THE COMPARISON WITH OTHER METHOD

Algorithm	**Recognition Probability**	**Average Recognition Time (msec)**
This Study	94%	4.54
Shan[11]	56%	2.57

VI. CONCLUSIONS

In this paper, we proposed a dynamic gesture recognition method based on fuzzy features, fuzzy neural network, and fuzzy inference system. First, it roughly classifies the dynamic gestures to circular gestures and linear gestures. Further, it sub-classifies the gestures to up, down, left, right, clockwise, and counter-clockwise. The experimental results show that this method has good recognition performance, makes the system to achieve better fault tolerance, and is more applicable to real gesture-controlled human-computer interactive environment.

ACKNOWLEDGMENT

The authors are grateful for the supporting by the National Science Council of the Republic of China [grant numbers NSC 101-2220-E-008 -002] and Funding from the Joint Research Center of National Central University and Delta Electronics Inc., under the project NCU-DEL-101-A-08.

REFERENCES

[1] C. Manresa, J. Varona, R. Mas and F. Perales, 2005, "Hand tracking and gesture recognition for human-computer interaction", ELCVIA, Volume 5, No. 3, p.96-104

[2] C. Manresa, J. Varona, R. Mas and F. Perales, 2005, "Hand tracking and gesture recognition for human-computer interaction", ELCVIA, Volume 5, No. 3, p.96-104

[3] W. T. Freeman and M. Roth, 1995, " Orientation Histograms for Hand Gesture Recognition", IEEE Intl. Wkshp. on Automatic Face and Gesture Recognition, Volume 50, Issue 2, pp.174.

[4] K. J. Chang, 2005, "Computer Vision Based Hand Gesture Recognition System",Master Thesis, National Tsing Hua University, Department of Electrical Engineering, page 21-32 .

[5] L.K. Chang, 2002, "Hand Gesture Recognition Based on Hausdorff Distance", Journal of Image and Graphics, Volume 7, No.11.

[6] M. Elmezain, A. Al-Hamadi, B. Michaelis, 2009, "Hand Locus-based Gesture Spotting and Recognition Using HMM", ICIP 16th IEEE International Conference , pp. 3577-3580.

[7] K. Murakami and H. Taguchi, 1991, "Gesture Recognition using Recurrent Neural Networks", Proc. of the SIGCHI conference on Human factors in computing systems, pp.237-242.

[8] Kosko, Bart (1992). Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Englewood Cliffs, NJ: Prentice Hall. ISBN 0-13-611435-0.

[9] F. J. Lin, W. J. Hwang, and R. J. Wai, 1999, "A supervisory fuzzy neural network control system for tracking periodic inputs," IEEE Trans. Fuzzy Systems, Volume 7, No.1, pp. 41-52.

[10] Y. C. Chen and C. C. Teng, 1995, "A model reference control structure using a fuzzy neural network," Fuzzy Sets and Systems, Volume 73, pp.291-312.

[11] C. Shan, T. Tan,Y. Wei, 2007, "Real-time hand tracking using a mean shift embedded particle filter", Pattern Recognition, Volume 40, Issue 7, pp.1958-1970.

[12] A. Bobick, J. Davis, 2001, "The recognition of human movement using temporal templates", IEEE Trans. Pattern Anal. Mach. Intell., Volume 23, No.3, pp.257-267.

BrainBrush, a Multimodal Application for Creative Expressivity

Bram van de Laar, Ivo Brugman, Femke Nijboer, Mannes Poel and Anton Nijholt
Human Media Interaction, Faculty of EEMCS
University of Twente
Enschede, The Netherlands
laar@ewi.utwente.nl

***Abstract*—We combined the new developments of multimodal Brain-Computer Interfaces (BCI) and wireless EEG headsets with art by creating BrainBrush. Users can paint on a virtual canvas by moving their heads, blinking their eyes and performing selections using a P300 BCI. A qualitative evaluation (n=13) was done. A questionnaire was administered and structured interviews were conducted to assess the usability and user experience of the system. Most participants were able to achieve good control over the modalities and able to express themselves creatively. The user experience of the modalities varied. The use of head movement was considered most positive with the use of eye blinks coming in second. Users were less positive about the use of the BCI because of the low reliability and higher relative cost of an error. Even though the reliability of the BCI was low, the BCI was considered to have an added value: the use of BCI was considered to be fun and interesting.**

***Keywords*-Creative Expression, Brain-Computer Interface, Multimodal Interaction, P300**

I. Introduction

The act of creative expression is considered by many to be a purely human ability and skill [1]. Creative expression allows humans to express their identity and it can take multiple forms: for instance making music, dancing, painting, writing or acting. Apart from these more traditional forms of art, new art forms have emerged, such as computer art, motion graphics and the use of virtual reality environments for art.

In recent years, a neurorevolution has taken place in which neurotechnology is a hot topic. For instance, research has been done on improving task performance using neurotechnology. Some athletes now use neurofeedback training to enhance their performance by managing the stress of training and competition [2]. Furthermore, Ros et al. have shown that the microsurgical skills of ophthalmic microsurgeons can be improved significantly with neurofeedback training [3]. Neurotechnology has even been applied to the field of economics, creating the science of neuroeconomics: combining neuroscience, economics and psychology to explain the human decision making process [4]. In neuromarketing, neurotechnology is used to research why consumers make certain choices: for instance, why consumers prefer either Pepsi or Coca Cola [5]

Brain-Computer Interfaces (BCIs) are another example of research in the field of neuroscience are. A BCI provides a direct interface between a human brain and a computer, without using peripheral nerves or muscles. For years BCI research has mainly focused on assistive technology for people with disabilities. For instance, patients with the motor neuron disease Amyotropic Lateral Sclerosis (ALS) can now benefit from BCIs like the P300 speller [6], brain-controlled wheelchairs [7] or BCIs to control their environment [8].

In recent years, BCI research has also focused on applications for healthy people. One example is the 'NeuroPhone' system developed by Campbell et al. which allows neural signals to drive mobile phone applications on an iPhone using a wireless EEG headset, for instance to dial a phone number using the same principles as the P300 speller [9].

Furthermore, research is being done on incorporating BCI technology in games. Plass-Oude Bos et al. developed AlphaWoW, incorporating BCI in the game 'World of Warcraft®' [10]. They used alpha waves in the EEG for automatic adaptation of the avatar shape from bear to elf and vice versa. Gürkök et al. used SSVEP for sheep herding in the game 'Mind the Sheep!' [11].

BCIs can also provide a unique link between the source of creativity, the brain, and art. Christoph De Boeck created a responsive environment, Staalhemel (http://www.staalhemel.com), where 80 steel segments are suspended in a room, above the visitor's head. The visitors wear a portable EEG headset and as they walk through the room, tiny hammers are activated by their brainwaves, tapping rhythmical patterns on the steel segments.

Other examples of the use of Brain-Computer Interfaces for art include the 'Brain-Computer-Music Interface' which enables a disabled person to create music [12], and the 'Câmara Neuronal', a performance where the brain signals of the performer are translated into audio and visual compositions (http://projects.jmartinho.net/3486412/Camara-Neuronal-Video-Teaser).

The German artist Adi Hösle, in cooperation with the Institute of Medical Psychology and Behavioural Neurobiology at the University of Tübingen, designed the application 'Brain Painting', a painting application which is controlled using a BCI and enables paralyzed patients to express themselves creatively. In 'Brain Painting', all actions are

 ISBN: 978-1-61208-250-9

performed using the P300 paradigm. The system uses two screens for the painter: one screen displays the P300 matrix while another, larger, screen shows the painting canvas. The standard P300 speller matrix, containing characters and numbers, was adapted to contain symbols indicating different colors, objects, grid sizes, object sizes, transparency, zoom and cursor movement. By repeatedly making selections using this P300 speller matrix, users can paint pictures on the virtual canvas. In the first evaluation of the 'Brain Painting' application, with 3 ALS-patients and 10 healthy subjects, both the ALS-patients and the healthy subjects were able to use the application with high accuracies: during a copy-painting task, the ALS patients achieved an average copy-painting accuracy of 70.18% while the healthy subjects scored an average accuracy of 80.53% [13]. One participant in the 'Brain Painting' study, who was severely disabled due to ALS, described her experience with the system: *"I am deeply moved to tears. I have not been able to paint for more than 5 years. Today I again had butterflies in my stomach, a feeling that I have missed for so much, so much* [sic]. *I was so sad, I was plagued by fears of loss, I was in shock because I could not paint. For me the picture I have created is so typical for me, no other paints in my style, and despite five years of absence, I am simply an artist again; I'm back to life!"*. Even though this feedback is very positive, the artistic freedom of the painter is limited due to the fact the cursor cannot be moved freely; the cursor can only be moved in a predetermined grid.

In continuation of the 'Brain Painting' research, Holz et al. developed the 'Brain Drawing' application to overcome the cursor movement limitation of the 'Brain Painting' system. In the 'Brain Drawing' application, imagined movement is used to control the cursor when drawing. During the first evaluation with 1 subject, the subject performed a Copy Drawing task in which he was instructed to draw a simple object (circle, ellipse or rectangle) on a virtual canvas. Holz et al. considered 4 out of 36 copied drawings to be succesful by visual inspection and the subject found it very difficult to draw. The subject had to focus his attention for a long period because he continuously had to imagine movement, which resulted in high workload [14].

Todd et al. used two different BCIs in their research on how creativity can be supported and assessed using a BCI [15]. With their first BCI, users could only control a drawing cursor in horizontal and vertical directions by looking at one of four LEDs placed at the top, bottom, left and right of the screen. The cursor would move in the direction of the LED the user looked at, and continue drawing in that direction until the user looked at another LED. For the second application, the four LEDs were mapped onto four shapes (circle, star, square and line). After choosing a shape, the shape would be drawn on the canvas. Users did not have any control over the position, size or color of the shapes. Todd et al. concluded that relying completely on the efficiency of a BCI for image production is not practical as BCI technology is not yet mature enough for 100% reliability. A possible solution they suggest is to create a hybrid, or multimodal, BCI by combining a BCI with other input modalities such as an eye-tracker.

These examples show that the focus in BCI research should shift from reliability to usability and user experience as is also reported by the FutureBNCI roadmap [1]. This shift in focus is necessary in order for BCIs to migrate out of the lab and into society. Healthy persons can choose from various alternative input modalities. For healthy persons to choose for BCI, the user experience and usability must be adequate. Most people have never used a BCI and the novelty of this new technology can be a reason for people to decide to use a BCI instead of alternative input modalities, even if a BCI is less reliable and slower. However, if the user experience and usability are not good, people are expected to choose a different input modality which provides a better user experience and usability. Due to the fact that the focus in BCI research has mainly been on the reliability, no standardized methods to assess the user experience for BCIs exist yet. Gürkök et al., Plass-Oude Bos et al. and Van de Laar et al. addressed the need for standardized methods to assess the user experience for BCIs [16], [17], [18]. Van de Laar et al. proposed a questionnaire consisting of a core containing general questions and modules for the different kinds of mental tasks and ways of interacting with the BCI [19].

In this study, we combined the new developments of multimodal BCIs and wireless EEG headsets with art by creating a multimodal interactive system (called BrainBrush) which allows healthy persons, but possibly also patients, to express themselves creatively. We evaluated this system to explore how the different modalities contribute to the user experience and whether BCI has an added value for this system.

II. Methods

First we will describe how BrainBrush was developed and how the design choices were made. Second we will outline how we set up our experiment to evaluate the system.

A. Development of BrainBrush

BrainBrush (see figure 1) was developed in a two step iterative process. The outcome of the first evaluation and how this influenced our design choices is however beyond the scope of this paper. For a more in-depth description see the work by Brugman [20].

We aimed to design the BrainBrush system in such a way that it would be appealing to healthy persons and would also be usable for patients who had lost control from their neck down (e.g. due to spinal cord injury). We expected healthy persons would find it appealing to be able to operate the BrainBrush system using only head movement, eyeblinks

 ISBN: 978-1-61208-250-9

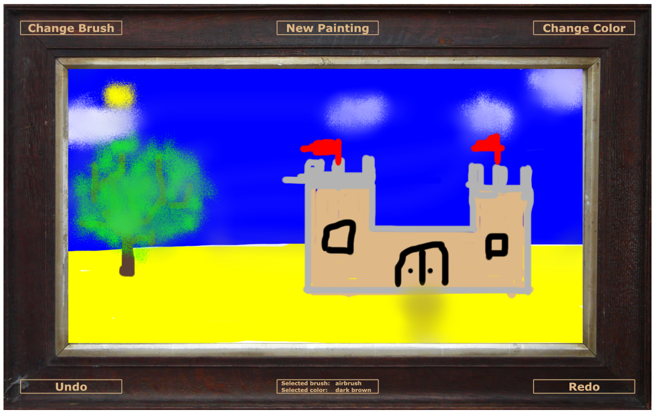


Figure 1. User interface in Painting mode with the ‘New Painting’ menu option at the top and the ‘Undo’ and ‘Redo’ menu options at the bottom

Figure 2. User interface during brush selection showing the 11 available brushes and the eraser

and brain activity because it is completely different to how healthy people normally use a computer. For patients who do not have control over their limbs, it is a necessity to be able to operate the system using nothing else than the described modalities.

Head movement, the P300 speller and eyeblinks were used as the input modalities for the BrainBrush system. Head movement has been used successfully in the past as an input modality, for instance to control a cursor [21]. Furthermore, the P300 speller has proven to be a robust BCI paradigm [22] and has been used in many BCI systems. The P300 signal is a positive deflection in the ongoing EEG signal observed roughly 300ms post stimulus over the centro-parietal area of the skull [6]. Finally, eyeblinks have been used for communication systems for ALS patients, such as the system for making selections on a computer screen proposed by Takeshita et al. [23]. However, the eyeblinks are usually detected using a camera and image processing techniques, instead of using an EEG headset. Chambayil et al. have shown promising results for their virtual keyboard BCI which uses eyeblinks to select characters [24]. All three modalities can be captured by one device: the Emotiv EPOC, using the EEG sensors and built-in gyro sensors.

The task of creating a brushstroke requires the system to provide users with a way to signal when they want the paintbrush to be put to the canvas and to signal once again when they want to take the paintbrush off the canvas, thereby ending the brushstroke. This on/off switch for the paintbrush has been implemented by detecting intentional eyeblinks in the EEG data based on an algorithm by Plass-Oude Bos et al. [25]. A template was constructed out of an EEG recording with 40 intentional eye blinks. A threshold is set for the Euclidean distance between the template and the live EEG data over channels AF3 and AF4. Both template and window have a length of 103 samples. When the Euclidean distance is below the threshold on either of the channels the on/off switch is triggered.

Head movement was implemented by using the gyro sensors of the Emotiv EPOC headset. An application called the *Mouse Emulator* is provided with the device and was used to convert head movement into cursor movement in order to move the paintbrush across the virtual canvas.

In BrainBrush, we use P300 speller grids for the selection of brushes and colors because the P300 speller is suitable for making selections from a large set of options in a relatively short period of time [26], compared to other Brain-Computer Interfacing paradigms. In a study performed by Guger et al. with healthy participants, 88.9% of the participants were able to achieve at least 80% accuracy using the P300 speller paradigm [22], showing the robustness of the P300 signal. Although, we suspect a lower accuracy in our system because of the lower signal quality of the EPOC hardware. For BrainBrush, we use the original grid structure, but instead of grids with characters, grids with pictures are used where each picture depicts a certain brush or color (see Figures 2 and 3). The *P3Speller* module of the BCI2000 framework is used for the implementation of the P300 speller grids [27]. The stimulus duration is set to 100ms and the inter-stimulus interval to 175ms. For the selection of a brush or color, 15 sequences of flashes are shown, meaning each row and column flashes 15 times. Therefore, the target is flashed 30 times in total.

Finally, BrainBrush uses the open source MyPaint application to provide us with a virtual canvas (see figure 1). An overview of the complete system can be seen in figure 4.

B. Experimental Design

The BrainBrush system was evaluated on user experience with a qualitative user study. We want to gain an understanding of underlying reasons and motivations for both the positive and negative opinions the participants formed about the user experience with the system and the various input modalities. Participants were asked to fill out an informed consent form and provide their demographic details. They

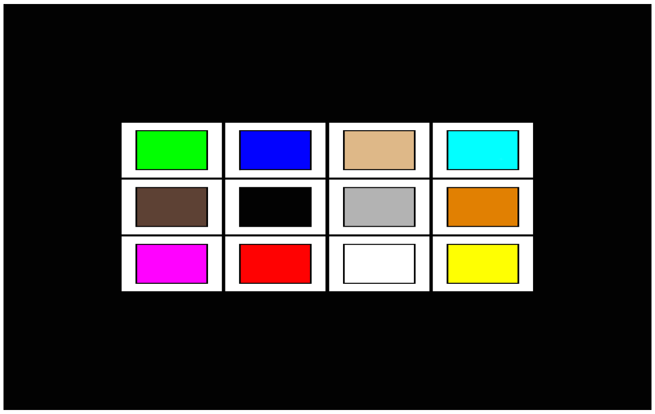


Figure 3. User interface during color selection showing the 12 available colors

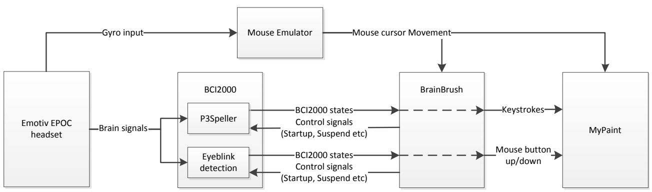


Figure 4. BrainBrush system overview

were given instructions on how to use the P3 Speller. They could try the P300 speller in a practice session, after any questions the experimenter proceeded to record a training session for the P300 classifier [28]. Training data consisted of a 10 letter word (‘BRAINPOWER’) with 15 sequences utilizing all 14 channels of the Emotiv EPOC. After the training session an online session was carried out in which participants tried to spell the 5 letter word ‘PAINT’ using the *P3Speller* without feedback so not to bias their expectations. Next, participants were given instructions on using BrainBrush. After 5 minutes of familiarizing themselves the free painting sessions started and the experimenter left the room. Participants could take as long as they wanted, up to 30 minutes. If they wanted to stop painting at an earlier point in time, they would ring a bell to signal the experimenter. After the free painting session, the experimenter would administer the SUS questionnaire [29] followed by an in-depth structured interview.

III. RESULTS

The group of participants for the user study consisted of 8 males (61.5%) and 5 females (38.5%). All participants were aged 20 to 29, the average age was 24.8 (standard deviation: 2.9). 11 participants (84.6%) are right-handed, 2 left-handed (15.4%). All participants had the Dutch nationality.

None of the participants reported relevant medical conditions. Out of 13 participants, 2 participants (15.4%) indicated that they had previously participated in BCI research. Four participants (30.8%) indicated they regularly exercise some

Table I
P300 CLASSIFICATION ACCURACY DURING COPY SPELLING AND FREE PAINTING

	Copy spelling		BrainBrush
Participant	Result	Accuracy	Accuracy
001	R6INW	40%	Regularly incorrect
002	XAINT	80%	70%
003	PACNZ	60%	80%
004	RHINT	60%	Regularly incorrect
005	R5IZT	40%	Regularly incorrect
006		-	< 50%
007	LAINT	80%	At the beginning: 50%
			At the end: 100%
008	RAIPT	60%	70%
009	RAINT	80%	21 out of 24 sessions correct (88%)
010	RAINT	80%	70%
011		-	Brush selection: 50-60%
			Color selection: 90%
012	L5BNN	20%	Choice decided beforehand: 100%
			Otherwise: lower
013	RGUNC	20%	25-33%
Average		56.4%	

form of creative expression: participant 001 paints, draws and designs daily for study-related purposes; participant 012 paints once a week as a form of recreation; participant 008 regularly uses Photoshop and participant 009 sometimes plays the drawing-game ‘Draw Someting’ on the iPad.

During the interviews after the free painting session, the participants were asked what accuracy they thought they had been able to achieve with the brush and color selection during the free painting session. These results together with the results from the online session are shown in table I.

During the experiments of participants 006 and 011, the classifier was accidentally not loaded during the online spelling session. Therefore, no copy spelling results are available for these participants. For the other 11 participants, the average accuracy during copy spelling was 56.4% (standard deviation: 23.4%, minimum 20%, maximum 80%).

The average SUS score for all participants was 66.2 (standard deviation: 14.2, minimum: 45, maximum: 87.5). A score of 68 would be average for all systems tested with SUS. During the interviews participants were asked to order the three modalities on pleasantness. An overview of the results can be seen in figure 5.

A summary of all analyzed topics covered in the questionnaire can be seen in table II.

IV. DISCUSSION AND CONCLUSION

In this study, we set out to develop a multimodal interactive system which allows persons to express themselves creatively and included Brain-Computer Interfacing technology. Using this multimodal interactive system, we wanted to research how the different modalities contribute to the user experience and whether BCI has an added value.

We developed the BrainBrush system, which lets users paint on a virtual canvas using their head movement for brush control, eyeblinks to turn the brush on and off, and

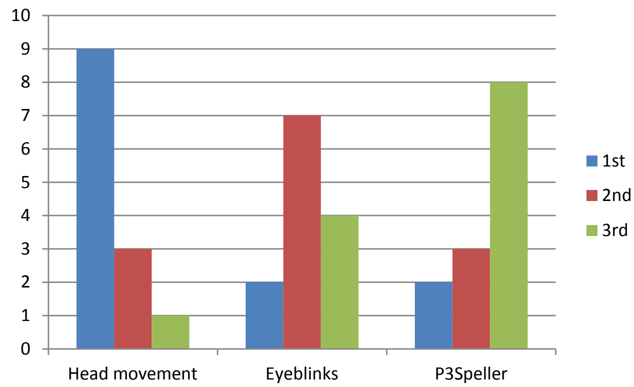


Figure 5. Ordering of pleasantness of modalities: Head movement was most often ranked most pleasant, eyeblinks most often second and the P3Speller (BCI) most often last.

Table II
SUMMARY OF INTERVIEW RESULTS

	+	+/-	-	Total
Expectations	6	7	0	13
Transfer creative ideas	4	3	6	13
Draw picture in mind	3	7	3	13
Time spent painting	4	2	7	13
Purchasing the system	0	4	9	13
Eyeblinks	7	2	4	13
Head movement	7	4	2	13
P3Speller	4	3	6	13
BCI for creative expression	9	2	2	13
Combination of modalities	9	1	3	13
Fun	12	1	0	13

the P300 speller to select different brushes and colors. The presented system is the first of its kind utilizing these modalities.

A user study with thirteen participants showed that the BrainBrush system does not enable all users to express themselves creatively. The group of users who were able to achieve good control over all three input modalities were able to expres themselves creatively and made nice paintings. However, for all users to be able to achieve this, the reliability of the input modalities must be improved.

The head movement modality was considered to be the most pleasant. However, misalignment of the cursor and a lack of smoothness negatively influenced the user experience. Replacing the Emotiv Mouse Emulator program with new software to translate head movement to cursor movement, is expected to improve the user experience.

After the head movement modality, the use of eyeblinks to turn the brush on or off was considered to be the most pleasant. In general, the user experience for this input modality was positive. However, the user experience can be further enhanced by performing more research into the topic of intentional eyeblink detection and improving the detection.

The P300 speller was considered to be the least pleasant input modality. In contrast to the other two input modalities, the user experience was not positive: using the P300 speller was considered to be mentally tiring and it caused physical discomfort and frustration. Participants suggested improvements to the P300 training setup and to the BrainBrush system.

The multimodal aspect of the system was good for the user experience. The combination of the three input modalities within the BrainBrush system was considered to be positive: it was a combination that made sense to most users and the input modalities were well balanced.

Finally, concerning the value of BCI for a multimodal interactive system for creative expression, we can conclude that BCI does have an added value: even though there were some issues due to the BCI modality, the use of BCI was considered to be fun, cool and interesting.

We feel the BrainBrush system in its current state offers a good basis, and with the suggested improvements, the user experience should improve further.

V. ACKNOWLEDGEMENTS

This manuscript is largely based on Ivo Brugman's Master's thesis [20].

REFERENCES

[1] "Future BNCI: A Roadmap for Future Directions in Brain / Neuronal Computer Interaction Research," 2012. [Online]. Available: http://future-bnci.org/images/stories/Future_BNCI_Roadmap.pdf

[2] M. Dupee and P. Werthner, "Managing the stress response: The use of biofeedback and neurofeedback with olympic athletes," *Biofeedback: Fall 2011*, vol. 39, no. 3, pp. 92–94, 2011.

[3] T. Ros, M. Moseley, P. Bloom, L. Benjamin, L. Parkinson, and J. Gruzelier, "Optimizing microsurgical skills with eeg neurofeedback," *BMC Neuroscience*, vol. 10, no. 1, p. 87, 2009.

[4] A. Rustichini, "Neuroeconomics: what have we found, and what should we search for," *Current Opinion in Neurobiology*, vol. 19, no. 6, pp. 672 – 677, 2009.

[5] S. M. McClure, J. Li, D. Tomlin, K. S. Cypert, L. M. Montague, and P. Montague, "Neural correlates of behavioral preference for culturally familiar drinks," *Neuron*, vol. 44, no. 2, pp. 379 – 387, 2004.

[6] L. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510 – 523, 1988.

[7] F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. d. R. Millán, "A Brain-Actuated Wheelchair: Asynchronous and Non-Invasive Brain-Computer Interfaces for Continuous Control of Robots," *Clinical Neurophysiology*, vol. 119, no. 9, pp. 2159–2169, 2008.

[8] U. Hoffmann, J. M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based braincomputer interface for disabled subjects," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 115 – 125, 2008.

[9] A. Campbell, T. Choudhury, S. Hu, H. Lu, M. K. Mukerjee, M. Rabbi, and R. D. Raizada, "NeuroPhone: brain-mobile phone interface using a wireless EEG headset," in *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds*, ser. MobiHeld '10. New York, NY, USA: ACM, 2010, pp. 3–8.

[10] D. Plass-Oude Bos, B. Reuderink, B. L. A. van de Laar, H. Gürkök, C. Mühl, M. Poel, A. Nijholt, and D. K. J. Heylen, "Brain-computer interfacing and games," in *Brain-Computer Interfaces. Applying our Minds to Human-Computer Interaction*, ser. Human-Computer Interaction Series, D. Tan and A. Nijholt, Eds. London: Springer Verlag, July 2010, pp. 149–178.

[11] H. Gürkök, G. Hakvoort, and M. Poel, "Evaluating user experience with respect to user expectations in brain-computer interface games," in *Proceedings of the 5th International Brain-Computer Interface Conference, BCI 2011, Graz, Austria*, G. R. Müller-Putz, R. Scherer, M. Billinger, A. Kreilinger, V. Kaiser, and C. Neuper, Eds. Graz, Austria: Verlag der Technischen Universität Graz, September 2011, pp. 348–351.

[12] E. R. Miranda, W. L. Magee, J. J. Wilson, J. Eaton, and R. Palaniappan, "Brain-Computer Music Interfacing (BCMI) From Basic Research to the Real World of Special Needs," *Music and Medicine*, vol. 3, no. 3, pp. 134–140, 2011.

[13] J. Münßinger, S. Halder, S. Kleih, A. Furdea, V. Raco, A. Hösle, and A. Kübler, "Brain Painting: First Evaluation of a New BrainComputer Interface Application with ALS-Patients and Healthy Volunteers," *Front Neurosci*, vol. 4:182. doi: 10.3389/fnins.2010.00182, 2010.

[14] E. M. Holz, T. Kaufmann, D. Franz, A. Hösle, and A. Kübler, "Brain Drawing: First Evaluation Results," in *Proceedings of the 3rd TOBI Workshop*, Würzburg, Germany, March 2012, pp. 109–110.

[15] D. A. Todd, P. J. McCullagh, M. D. Mulvenna, and G. Lightbody, "Investigating the use of brain-computer interaction to facilitate creativity," in *Proceedings of the 3rd Augmented Human International Conference*, ser. AH '12. New York, NY, USA: ACM, 2012, pp. 19:1–19:8.

[16] H. Gürkök, D. Plass-Oude Bos, B. L. A. van de Laar, F. Nijboer, and A. Nijholt, "User Experience Evaluation in BCI: Filling the Gap," *International Journal of Bioelectromagnetism*, vol. 13, no. 1, pp. 54–55, July 2011.

[17] D. Plass-Oude Bos, H. Gürkök, B. L. A. van de Laar, F. Nijboer, and A. Nijholt, "User Experience Evaluation in BCI: Mind the Gap!" *International Journal of Bioelectromagnetism*, vol. 13, no. 1, pp. 48–49, July 2011.

[18] B. van de Laar, F. Nijboer, H. Gürkök, D. Plass-Oude Bos, and A. Nijholt, "User Experience Evaluation in BCI: Bridge the Gap," *International Journal of Bioelectromagnetism*, vol. 13, no. 3, pp. 157–158, 2011.

[19] B. van de Laar, H. Gürkök, D. Plass-Oude Bos, F. Nijboer, and A. Nijholt, "Perspectives on user experience evaluation of brain-computer interfaces," in *Universal Access in Human-Computer Interaction. Users Diversity*, ser. Lecture Notes in Computer Science, C. Stephanidis, Ed. Springer Berlin / Heidelberg, 2011, vol. 6766, pp. 600–609.

[20] I. Brugman, "A Multimodal Interactive System for Creative Expressions," August 2012. [Online]. Available: http://essay.utwente.nl/61988/

[21] G. Evans and P. Blenkhorn, "A head operated joystick – experience with use," in *Proceedings of the CSUN conference on technology and persons with disabilities*, California State University, Northridge, 1999.

[22] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger, ""how many people are able to control a p300-based braincomputer interface (bci)?"," *Neuroscience Letters*, vol. 462, no. 1, pp. 94 – 98, 2009.

[23] K. Takeshita, A. Uchibori, Y. Mizukami, T. Satoh, K. Tanaka, and S. Uchikado, "A communication system for ALS patients using eye blink," *International Journal of Applied Electromagnetics and Mechanics*, vol. 18, no. 1-3, pp. 3–10, 2003.

[24] B. Chambayil, R. Singla, and R. Jha, "Virtual keyboard BCI using Eye blinks in EEG," in *Wireless and Mobile Computing, Networking and Communications (WiMob), 2010 IEEE 6th International Conference on*, oct. 2010, pp. 466 –470.

[25] D. Plass-Oude Bos, M. Duvinage, O. Oktay, J. Delgado Saa, H. Guruler, A. Istanbullu, M. van Vliet, B. L. A. van de Laar, M. Poel, L. Roijendijk, L. Tonin, A. Bahramisharif, and B. Reuderink, "Looking around with your brain in a virtual world," in *Proceedings of the 6th International Summer Workshop on Multimodal Interfaces, eNTERFACE'10*. University of Amsterdam, 2010, pp. 12–23.

[26] F. Aloise, P. Aricò, F. Schettini, E. Lucano, S. Salinari, F. Babiloni, D. Mattia, and F. Cincotti, "Can the P300-based BCItraining affect the ERPs?" *International Journal of Bioelectromagnetism*, vol. 13, no. 3, pp. 148–149, 2011.

[27] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1034 –1043, june 2004.

[28] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the P300 Speller," *Journal of Neural Engineering*, vol. 3, no. 4, p. 299, 2006.

[29] J. Brooke, "SUS - A quick and dirty usability scale," in *Usability Evaluation In Industry*, P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester, Eds. CRC Press, 1996, pp. 189–194.

Multimodal Human-Robot Interactions: the Neurorehabilitation of Severe Autistic Children

Irini Giannopulu
Cognitive Neuroscience
Medical Engineering and Reeducation
Pierre & Marie Curie University
email: igiannopulu@gmail.com

***Abstract*—In the context of human-robot interactions, we studied quantitatively and qualitatively the interaction between autistic children and a mobile toy robot during free spontaneous game play. A range of cognitive nonverbal criteria including eye contact, touch, manipulation, and posture were analysed, firstly in a dyadic interaction and secondly in a triadic interaction. Once the cognitive state between the child and the robot established, the child interacts with a third person displaying positive emotion. Both dyadic and triadic interactions of autistic children and a mobile toy robot suggest that the mobile toy robot in an ecological situation such as free, spontaneous game play could be used as a neural mediator in order to improve children's brain activity**.

Keywords-multimodal interactions ; severe autism ; mobile toy robot ; spontaneous free game play ; neural mediator.

I. Introduction

Different kinds of computer based technologies such as robots and virtual reality, i.e., human-machine-interactions (HMI), are being put to effective use in the education of autistic children. The studies we developed aimed to analyse the multimodal interactions of severely autistic children, during free game play with a mobile toy robot in both dyadic and triadic situations.

Autism is a severe neurocognitive disorder. Because the effects of autism can range from severe to mild, autism is considered to be a spectrum disorder [1]. Severe autism is characterized by repetitive and stereotypical behavior apparent by 3 years of age, impairment in verbal and nonverbal processes, emotional and social interaction. Its genetic and neurocognitive aetiology is unknown; different hypotheses have been and continue to be discussed. Autism is considered a complex multifactor disorder involving many genes [2], [3], [4]. These findings have given rise to new insights into neuronal circuits relevant to autism disorders. As would be expected, a large number of functional neuroimaging studies have demonstrated that different brain regions are involved in autism. In particular, these studies show that the neural substrate underlying cognitive, social and emotional impairment involves multimodal areas such as the exterior superior temporal sulcus [5], the interior temporal lobe, amygdala included [6], as well as the ventral part of the prefrontal cortex, i.e., orbital frontal cortex [7]. In addition, autistic children also show aberrant brain connectivity and disruption of white matter tracts between temporal regions [8] which disrupt verbal and nonverbal acquisition, consolidation as well as social interaction [9], [10], [11], [12]. Taken together, the aforementioned studies provide the basis for concluding that in autism the more impaired cortical areas are those that are involved in complex cognitive functions such as perception, language, social interaction and emotion. Such complex expression of autism necessitates a more generic consideration of this disorder at the neural level.

From a developmental viewpoint, the most widely accepted hypothesis in autism is the theory-of-mind deficit [13]. Even if this theory cannot account for the whole spectrum of autistic disorders, it raises many issues which not only involve mental representation of others but also social skills such as posture [14], eye contact [15], touching [8] and manipulation [16] that express social interaction [17].

Game play is a very important feature of early childhood and is of particular importance for children with autism. Play in children with autism is more like "learned routine" rather than "spontaneous" [18]. Autistic children show difficulty in their play activities which could be associated with their deficit in cognitive, and emotional development. Free game play characterized by spontaneity could allow children with autism the possibility to express themselves and engage in satisfying social activity which in turn could lead to development of their cognitive skills.

Different approaches are currently being utilized to better understand the capacity of autistic children to interact with a robot [19]. The Aurora project investigates the use of robots (Labo-1, Kaspar, Robota doll, for example) in game play. The aim is to create a tool based on an autonomous robot that convinces autistic children to engage in a process of interaction [20], [21], [22] [23]. The interactions are tested through the analysis of visual contact, joint attention, avoidance or fleeing, visual pursuit, and whether the child imitates the robot [24]. Using a variety of modalities for interaction such as music, colour and visual contact, a sensitive robot named Tito was employed in social interaction with autistic children [21], [25]. A very small fixed robot named

 ISBN: 978-1-61208-250-9

Keepon can capture and maintain visual contact with the child, drawing its attention and initiating some element of conversation [21], [26]. Roboto uses the form of an animated face (mouth, eyebrows, eyes) that can cause behaviour imitation from the part of the autistic child [27].

Regardless of the child's mental age, all these studies have reported dyadic interaction between the autistic child and the robot. Even though (because of the pathology) the number of the children participating in these experiments is limited, the dyadic interaction is reflected in attention [22], imitation cognition [27], visual contact [28], touching and verbal conversation [25], manipulating and posture [22]. All these studies have shown that animated robots, humanoid or not, using different stimulation encourage interaction in autistic children. In other words, HMI, i.e., robot in our case, could be used to improve autistic children's behaviour. In the above studies, the focal point of the analysis was on a single mode of interaction. Even if quantitative metrics of social response for autism diagnosis including robots were developed [29]; only one study has used a quantitative technique for analysing dyadic interaction for autism therapy [30]. With the exception of Labo-1 in the Aurora project, and Roball in Michaud's project so far, only fixed robots have been utilized reducing the child's spontaneity and self–expression in game play.

We used a mobile toy robot named "GIPY 1" (Fig. 1) which incites the child to engage in interaction. On the hypothesis that autistic children will be in quasi-constant interaction with the robot, the cognitive behavior of autistic children in interactive activities with a robot, i.e., dyadic interaction, during spontaneous game play using multimodal criteria was analyzed. In addition, we hypothesized that once dyadic interaction is established, the child could use the robot as a mediator to initiate the interaction with the third person, an adult, and express emotion, i.e., triadic interaction. This cognitive and emotional interaction of the autistic child with a third person was investigated, once again, in spontaneous, free game play by means of a multimodal approach.

The structure of the paper is the following : first we will give the method for both dyadic and triadic interactions ; then, we will analyse the results for both interactions ; finally, we will develop the discussion, the conclusion and the future work.

II. Method

A. Participants

- Dyadic interaction

Four children (3 boys and 1 girl) participated in this study. Their chronological ages ranged from 7 to 9 years old (mean 8.3 years). Their developmental age ranged from 2 to 4 years old. The children were diagnosed according to the D.S.M. IV-TR criteria of autism [31]. The C.A.R.S [32] had been administered at the age of 6 years by an experienced clinical psychologist. The C.D.I [33] was used to estimate intellectual disability (Table 1). At the time of the experiment all of the children were attending special education classes or autism.

Table 1. General characteristics of population
a) Childhood b) International Classification of Diseases

Children	Chronological age	Sex	C.A.R.S (a)	C.D.I (b)
1	7y 11m	m	46.5	20 to 34
2	8y 6m	m	35.5	35 to 49
3	9y 5m	f	31.4	20 to 34
4	8y 2m	m	43.5	20 to 34

- *Triadic interaction : Case Study*

"A" is a right-handed young boy. He exhibits mental retardation as per the C.D.I. [33]. His chronological age is 8 years old and his developmental age is 2 years old. The child was diagnosed with autism when he was 3 years old and still displays all characteristics of autism according to the D.S.M IV-TR [31]. In addition, the C.A.R.S. [32] has shown severe autism with a score of 43 points. "A" has deficits in reciprocal social interactions and communication (speech and language), stereotyped behaviour and restricted interests and activities. At the time of the experiment he was attending special education classes for autism.

B. Material

- *Room*

The room was 4.56 m by 3.34 m. A chair, a small wardrobe and a table on which the equipment needed for the framework of the study was placed (laptop and joystick), were used. In order to reduce the presence of disruptive elements and so as to avoid autistic bend, the room was left bare [34].

- *Robot*

A mobile robot, called "GIPY-1", which is cylindrical-shaped with a diameter of 20 cm and a height of 30 cm, was created for use in the experiment. A representation of a neutral facial expression constitutes the cladding of the robot: the round eyes and nose triangle were dyed olive green and the elliptical mouth was dyed red (Fig. 1). Everything was covered with a transparent plastic sheet. The simplicity of the robot was driven by the preference of autistic children for simple and predictable toy design [35]. An operator manipulated the robot via a wireless remote control using a joystick connected to a laptop. The robot could move forward, backward and turn on itself at low speed. These movements were constant.

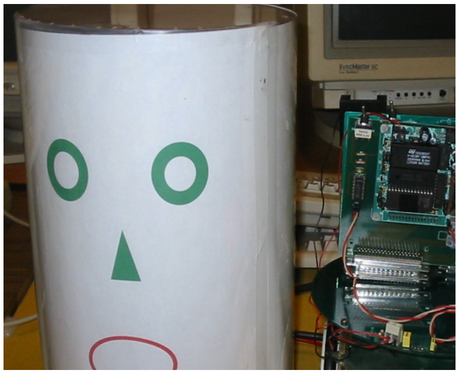


Figure 1. GIPY 1

- *Protocol for the dyadic and the triadic interactions*

The duration of the session was 5 minutes. The robot was placed on the ground beforehand, in the center of the room, its stylised face toward the entrance. The game play session began as follows: when the child and the adult entered the room, the tele-operated robot carried out three movements (move forward, move back, 360° swivel). As in real social interaction, the child and the robot altered their responses. If the child approached, the robot moved back and conversely. If the child moved away from the robot, i.e., ignored the robot, the robot followed the child in order to attract its attention. If the child remained motionless, the robot approached or turned itself around in order to focus the attention of the child. All movements were standardised.

- *Analysis for the dyadic and the triadic interactions*

Two independent judges unfamiliar with the aim of the study completed the observations of the game play skills. Both performed the analyses of video sequences with Elan software. Prior to assessing game play improvement, inter-judge reliability was assessed to ensure that both judges who analysed videotapes were consistent in their analyses. Inter-judge reliability was assessed using intra-class coefficients to make the comparison between them. The inter-judge reliability was good (Cohen's kappa=0.63).

The dependent variable was the time of child-robot interaction for the dyadic interaction and the time of child-robot and adult for the triadic interaction. Accordingly, we calculated the duration of all the characteristics of each criterion. This was defined as the duration between the onset time and the offset time of each child's behaviour toward the robot. Four criteria were defined for the dyadic interaction: 1) eye contact (looking at the robot), 2) touch (touching the robot without manipulating it), 3) manipulation (operating the robot), 4) posture (changing corporal position toward the robot). Based on the hypothesis that cognitive interaction could be lead to the expression of an emotional state, a n additional fifth criterion was defined for the triadic interaction. This criterion was: the positive emotion (display of enjoyment) (5). The duration of each criterion was calculated in seconds and was considered independent of the others. Concerning, for example, the characteristic "s/he looks at the immobile robot" ("eye contact") the onset time corresponded to the time when the child looked at the robot and the offset time to the moment when the child looked away from the robot. We calculated the duration of all the characteristics of each criterion. We summed up the duration corresponding to each criterion. Only the total duration is presented in the results section.

III. Results

- *Dyadic interaction*

The mean time of dyadic interaction was 238.7 sec. In other words, the children spent nearly 80% of their time (156 seconds for the first, 289 seconds for the second, 269 seconds for the third and 241 seconds for the forth child) playing with the robot. The duration of each robot-child interaction is presented in Fig. 2. The duration of "eye contact" is similar for all the children. However the analysis of the duration of "touching", "manipulating" and "posture" possibly reflects inter-individual differences related to different forms of autism. This analysis also showed how autistic children's behavioral interaction with the robot changes over a period of time. This suggests that a mobile toy robot could help autistic children to reduce repetitive and stereotypical behavior.

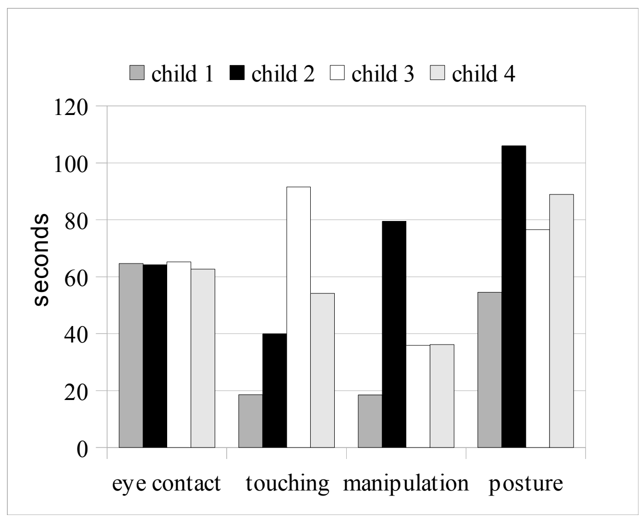


Figure 2. Duration of dyadic interaction for each criterion

- *Triadic interaction*

The mean time of dyadic interaction was 25 sec; the mean time of triadic interaction was 30 sec. In other words, the child spends half the time playing with the robot and the half the time playing with the robot and the adult.

 ISBN: 978-1-61208-250-9

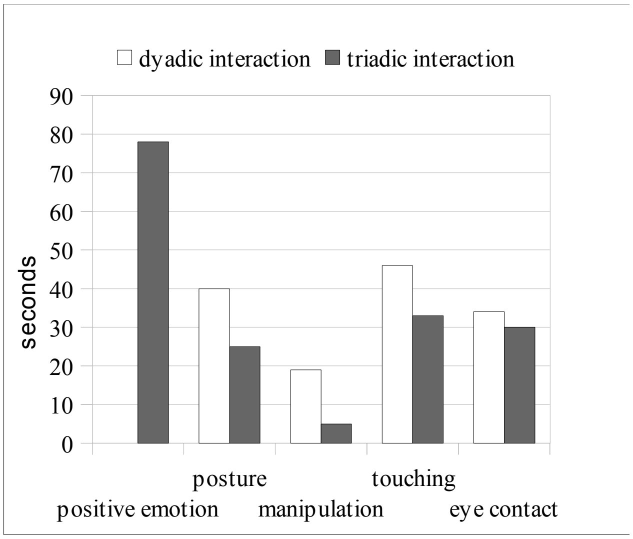


Figure 3. Duration of dyadic and triadic interactions for each criterion

The duration of dyadic and triadic interaction is presented in Fig. 3. The duration of "eye contact" and of "touching" is similar in both situations. However the duration of "manipulation", of "posture" and of "positive emotion" differ between the two situations. As we can observe, positive emotion is more easily expressed when the child interacts with the adult and the robot than when the child interacts only with the robot. This difference reflective of the changes in autistic child behavior with the robot over a period of time also tells us that a mobile robot could be used as a mediator for social and emotional interaction. This is an encouraging conclusion with regard to the potential of human-to-human interaction.

IV. DISCUSSION

- *Dyadic interaction*

The aim of the first study was to analyze the interaction between autistic children and a mobile toy robot in free, spontaneous game play. Consistent with our hypothesis, the children are quasi-constantly in interaction with the mobile robot using a variety of ways.

As autism is a spectrum disorder where a large variation in abilities and interests among autistic children is apparent, the interaction of children and robots was evaluated on the level of each individual child. Consistent with various studies, the present study shows that the use of robots engages autistic children in interaction [20], [24-27], [36-39]. We have calculated the duration of robot-child interaction during free, spontaneous game play.

More precisely, the behavior of autistic children vis-a-vis the robot based on four criteria ("eye contact", "touching", "manipulation" and "posture") has been analyzed and a temporal quantification of dyadic interaction with respect to the duration was performed. The analysis revealed that the duration of eye contact behavior was similar for each child. Inter-individual differences were identified for the duration of "touching", "manipulating" and "posture" behavior. These differences might be related to different forms of autism. This data demonstrated that the autistic children not only visually explored the robot [30] but also engaged in different kinds of play with the robot. In other words, the autistic children clearly took an interest in playing with the mobile robot.

In all the studies we have mentioned above, only fixed robots were used, with the exception of Labo-1 [22] and Roball [21]. In our study as with Labo-1 and Roball studies, the autistic children were invited to interact with the robot during free, spontaneous game play. Taken together, these studies have shown that autistic-children use a variety of behavior when playing with a robot in free game play.

It seems that free game play could be a relevant ecological situation where an autistic child spontaneously interacts with the robot. Moreover, mobile toy robot could help autistic children to reduce repetitive and stereotypical behavior.

These findings also reveal that free, spontaneous game play with robots is possible with severe autistic children and could better facilitate the transfer of social and learnt abilities to real life.

But what is important to demonstrate is whether and how autistic children can generalize learnt abilities during play with the robot to aduts, i.e., proving that the robot could be used as a neural mediator tool for the enrichment of child-human interaction.

- *Triadic interaction*

In this case study, we analysed the ingredients of child-robot two-pronged interaction and child-robot-adult three-pronged interaction. Consistent with our hypothesis, the child first establishes a relationship with the robot and then uses the robot as an "instrument" to initiate the interaction with the adult. At first glance, our results are compatible with recent findings according to which HMI i.e., presence of a robot, are more effective than other environments in allowing autistic children to express social interest towards the robot [22-23], [25], [28], [35], [40-41]. In these studies, researchers have used robots for treating autistic children. However, the relationship between robot and child has been studied solely based on the analysis of a single mode of interaction. Furthermore, the studies have been conducted using fixed robots. Our results go beyond these findings because we have demonstrated, as far we know for the first time, that in spontaneous, free game play, an autistic child uses the robot to interact with the adult and to express positive emotion. As such, on the one hand, we have shown that the dyadic interaction is based on a cognitive state and, on the other, that the child uses the robot as a mediator to express positive emotion playing with the adult.

More precisely, in our study, the interaction between robot and child was analysed using different criteria such as

 ISBN: 978-1-61208-250-9

eye contact (looking at the robot), manipulating (operating with the robot), touching (touching the robot without manipulating it), posture (changing postural position toward the robot). Consistent with our previous studies [42-45], we have demonstrated that visual, haptic, tactile perception and posture, i.e., multimodal perception, are on the basis of the interest the child displays towards the robot. This is because, in our approach (as in Quinn & Eimas approach [46]), perception and cognition are considered to be a single domain rather than two distinct entities. The criteria we have chosen are assumed to represent the state of the child's cognitive processes, as expressed by the interest the child exhibits towards the robot in spontaneous, free game play. As our second study has shown, once this state is established, the child develops a triadic relation i.e., with the robot and the adult, thereby displaying enjoyment, which is a positive emotion. The expression of positive emotion could be related to the emergence of a cognitive state, which is multimodal in our case. This expression appears when the child interacts with the adult using the robot. This is a very important finding when we consider that the subject of our case study "A" exhibited a score of 43 which corresponds to severe autism. Individuals with severe autism exhibit very limited social skills. They don't express emotions. They don't respond well to behavioural therapy and in fact tend to show few, if any, signs of improvement after such therapy is undertaken. However, as we showed in our study, "A" is in constant interaction with the robot, expressed by a multimodal cognitive state which, according to us, allows him to express positive emotion with the adult.

V. CONCLUSION AND FUTURE WORK

Considering the above studies, it would be fair to conclude that autism therapy using robots seems to be effective, safe and convenient. What is important is the "passage" from dyadic interaction to triadic interaction. Indeed, when "A" interacts with both the robot and the adult, he changes his behaviour. What causes this behavioural modification? We think that the robot as a mediator could bring about neurocognitive improvements to the autistic child. As the results have shown, the extent of that improvement seems to be smaller when the child interacts with the robot than when he interacts with the adult. We believe that the child's reactions to the robot are very important in establishing child interest and are of paramount importance in robot therapy. In fact, this dyadic interaction could be thought of as the building block from which the relationship among humans may be developed. Consistent with this interpretation may be the fact that positive emotion is expressed only when the child interacts with the adult via the robot. Positive emotion is quasi-absent when the child interacts with the mobile robot on a standalone basis. It seems thus reasonable to infer that the three-pronged interaction i.e., child-robot-adult could better facilitate the transfer of social and emotional abilities to real life.

Moreover, in both studies, the findings tell us that free game play, i.e., an ecological situation, encourages an autistic child to interact with the robot in a spontaneous manner and could reduce repetitive and stereotypical behavior. They also reveal that free, spontaneous game play with robots is possible with autistic children and could better facilitate the transfer of learnt abilities to real life.

One limitation of these studies is the small number of autistic children which makes impossible inferential analysis. Additional studies are required with typical and autistic children. Longitudinal follow-up of the same children is necessery to examine the efficiency of mobile robots in improving the neurocognitive skills of autistic children. This is what we're developing currently. In addition, with a new study we analyse the embedded multimodal nonverbal and verbal interactions between a mobile toy robot and autistic children using a new paradigme.

ACKNOWLEDGMENT

The author would like to thank all the children who participated to the studies, their parents as well as ANR's supporting. The author would also like to thank the reviewers for the thoughtful comments, which improved the manuscript.

REFERENCES

[1] L. Crane, S.E. Lind, and D.M. Bowler, "Remembering the past and imaging the future in autism spectrum disorder," Memory, in press.

[2] L. Jorde et al. "Complex segregation analysis of autism," Am. J. Hum. Genet. vol. 49, pp. 932-938, 1991.

[3] N.H. Sykes and J.A. Lamb, "Autism: The quest for the genes," Exp. Mol. Med. vol. 9, pp. 1-15, 2007.

[4] P. Szatmari et al. "Mapping autism risk loci using genetic linkage and chromosomal rearrangements," Nat. Genet. vol. 39, pp. 319-328, 2007.

[5] K.A. Pelphrey and E.J. Caster, "Charting the typical and atypical development of the social brain," Dev. Psychopathol. vol. 20, pp. 1081-1102, 2008.

[6] B.A. Corbett et al. "A functional and structural study of emotion and face processing in children with autism," Psych. Res. 30, 173, pp. 196-205, 2009.

[7] L. Brothers, "The social brain: a project for integrating primate behaviour and neurophysiology in a new domain," Concepts. Neurosc. vol. 1, pp. 27–51, 1990.

[8] U. Frith and C.D. Frith, "Development and neurophysiology of mentalizing," Philos. Trans. R. Soc. London. vol. 358, pp. 459-473, 2003.

[9] A. Saitovitch et al. "Social cognition and the superior temporal sulcus: Implications in autism," Rev. Neurol. (Paris), 2012 Sep 13. pii: S0035-3787(12)00902-2. doi: 10.1016/j.neurol.2012.07.017.

[10] R. Adolphs and D. Tranel, «Intact recognition of emotional prosody following amygdala damage," Neuropsychol. vol. 37, pp. 1285–1292, 1999.

[11] J.P. Aggleton, The Amygdala: A Functional Analysis. Oxford: University Press, 2000.

[12] B.M. Nacewicz et al. "Amygdala Volume and Nonverbal Social Impairment in Adolescent and Adult Males with Autism," Arch. Gen.

Psychiatry. vol. 63, pp. 1417-1428, 2006.
[13] U. Frith, J. Morton, and A.M. Leslie, "The cognitive basis of a biological disorder: autism," TIN14. pp. 433-438, 1991.
[14] A.M. Leslie, "Pretending and believing: issues in the theory of ToMM," Cognition, vol. 50, pp. 211-238, 1994.
[15] T. Reed and C. Peterson, «A comparative study of autistic subjects' performance at two levels of visual and cognition perspecctive taking," J. Autism. Dev. Disord. vol. 20, pp. 555-567, 1990.
[16] A. Escalona, T. Field, J. Nadel, and B. Lundy, "Brief report: Imitation effects on children with autism," J. Autism. Dev. Disord. vol. 32, pp. 141-144, 2002
[17] B.A. Taylor et al. "Manipulating establishing operations to promote initiations toward peers in children with autism," Res. Dev. Disabil. vol. 26, pp. 385-392, 2005.
[18] E. Williams, V. Reddy, and A. Costal, "Taking a closer look at functional play in children with autism," J. Autism. Dev. Disord. vol. 31, pp. 67-77, 2001.
[19] B. Scassellati, H. Admoni, and M. Mataric, "Robots for use in autism research," An. Rev. Biomed. Eng. vol. 14, pp. 275-294, 2012.
[20] M. Davis, B. Robins, K. Dautenhahn, C.L. Nehaniv, and S. Powell editors. "A comparison of interactive and robotic systems in therapy and education for children with autism," European Conference for the Advancement of Assistive Technology in Europe (AAATE'05), pp. 353-357, September 6-9 2005, IOS Press Lille.
[21] F. Michaud and S. Caron, "Roball, the rolling robot," Auto. Rob. vol. 12, pp. 211-222, 2002.
[22] K. Dautenhahn, "Socially intelligent robots: dimensions of human-robot interaction," Philos. Trans. R. Soc. vol. B.362, pp. 679-704, 2007.
[23] B. Robins, K. Dautenhahn, and K. Dubowski, "Robots as Isoalators or Mediators for children with autism ? A Cautionary Tale," Proc's of the AISB 05 Symposium on Robot Companions, pp. 82-88, 2005.
[24] A. Duquette, F. Michaud, and H. Mercier, "Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism," Auto. Rob. – Special Issue. vol. 2, pp. 147-157, 2004.
[25] F. Michaud, T. Salter, A. Duquette, and J.F. Laplante, "Perspectives on mobile robots used as tools for pediatric rehabilitation," Assist. Technol. vol. 19, pp. 14-19, 2007
[26] H, Kozima, C Nakagawa, and Y. Yasuda, «Children-robot interaction: a pilot study in autism therapy," Prog. Brain. Res. vol. 164, pp. 385-400, 2007.
[27] J. Nadel, A. Revel, P. Andry, and P. Gaussier, "Toward communication: first imitations in infants, low-functioning children with autism and robots," Inter St: Soc. Beh. Commun. Biol. Syst. vol. 5, pp. 45-54, 2004.
[28] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, «Robot programming by demonstration," in Handbook of Robotics, chapter 59, B. Siciliano, and O. Khatib, Eds Springer, 2008.
[29] B. Scassellati, "Quantitative metrics of social response for autism Diagnosis," IEEE International Conference on Intelligent Robots and Systems, pp1134-1138 vol.2, 2002.
[30] Dautenhahn, K and Werry, I, "A quantitative technique for analysing robot-human interactions," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1132-1138, 2002.
[31] DSM-IV-TR Manuel diagnostique et statistique des troubles mentaux, Paris, Editions Masson, 2003.
[32] E. Schopler, R.J. Reichler, R.F. De Vellis, and K. Daly, "Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS)," JADD10, pp. 91-103, 1980.
[33] International Classification of Diseases, Revision, 10, 1990.
[34] P. Planche E. Lemonnier, K. Moalic, C. Labous, and A. Lazartigues, "Les modalités de traitement de l'information chez les enfants autistes," Ann. Med. Psychol. vol. 160, pp. 559-564, 2002.
[35] J. Pedersen and J. Schelde, "Behavioral aspects of infantile autism: an ethological description", Eur. Child. Adolesc. Psy. vol. 6, pp. 96-100, 1997.
[36] K. Dautenhahn and I. Werry, "Towards interactive robots in autism therapy," P & CT, 121, pp. 1–35, 2004.
[37] H. Kozima and H. Yano, Designing a robot for contingency-detection game. Working Notes Workshop Robotic & Virtual Interactive Systems in Autism Therapy, 2001 27-28 September, U.K.Hatfield. University of Hertfordshire, 2001, Technical Report No 364.
[38] J. Martineau, S. Cochin, R. Magne, and C. Barthelemy, "Impaired cortical activation in autistic children: Is the mirror neuron system involved?" Int. J. Psychophysiol. vol. 68, pp. 35–40, 2008.
[39] T. Salter, I.P. Werry, and F. Michaud, "Going into the wild in child-robot interaction studies - Issues in social robotic development," Int. J. Robot. vol. 1, pp. 93-98, 2007.
[40] Y. Yokoyama, The Possibility of the Psychiatric Treatment with a Robot as an Intervention -From the Viewpoint of Animal Therapy-, *Proc. of Joint 1st International Conference on Soft Computing and Intelligent Systems and 3rd International Symposium on Advanced Intelligent Systems, paper number 23Q1-1, in CD-ROM Proc. 2002.*
[41] C. Kozima and Y. Yasuda, "Children-robot interaction: a pilot study in autism therapy," Prog. Brain. Res. vol. 164, pp. 385-400, 2007.
[42] I. Giannopulu and G. Pradel, Mobile toy robots can be used in autism therapy: an example of application. *IEEE Proceedings in the IROS 2009 paper number SuT8.pdf in the workshop CD Proc. 2009. 2009a*
[43] I. Giannopulu and G. Pradel, (2009b). Interactions multimodales en situation de jeu libre entre enfants autistes et un robot mobile. 7ème Journée Nationale de la recherche en Robotique, http://jnrr09.lms.sp2mi.univ poitiers.fr/IMG/pdf/Giannopulu_Pradel. Novembre, Domaine de La Grande Garenne 18330 Neuvy-sur-Barangeon.
[44] I. Giannopulu and G. Pradel, "Multimodal interactions in free game play of children with autism and a mobile robot," NeuroRehabilitation. vol. 27, pp. 305-311, 2010.
[45] I. Giannopulu, "Cognitive and emotional interactions between autistic child, mobile toy robot and therapist," Frontiers in Computational Neurosciences, 2011 doi:10.3389/conf.fncom.2011.52.00002.
[46] P.C. Quinn and P.D. Eimas, "The emergence of category representations during infancy: are separate perceptual and conceptual processes required?" J. Cogn. Dev. vol. 1, pp. 55-61, 2000.

Présentation conf Nice ACHI 13

Multimodal human-robot interactions: the neurorehabilitation of severe autistic children

My talk draws from my recent work which concerns the multimodal interactions in typically and atypically developing children.

From my viewpoint, multimodal interactions (which mean different informations coming from different sources) express the complex relationship between brain, mind and environment (natural/physical or artificial environment).

Multimodal interaction is the foundation of the brain.

The individual history of a neuron can be summarized as following: a neuron creates direct and/or indirect connections with other neurons (via synapses). Biological in nature, this operation seems to contribute to the formation of a multimodal neuron. This seems to affect the future development of brain areas. The individual history of a brain area depends on its direct and/or indirect connections with the contiguous cortical areas. This constitutes the neural environment also known as the 'natural or physical environment'.

 ISBN: 978-1-61208-250-9

In typically developing children the realization of a given neurocognitive task such as verbal or nonverbal, social and emotional is possible because of a wide cortical network. The neural environment.

In atypically developing children such as autistic children the emergence of social and emotional behavior interaction could be possible because of a mobile toy robot, an articificial environment.

In order to study this hypothesis, we created an artificial environment using a mobile toy robot and we analyzed multimodal interactions in free game play with autistic children and that robot in dyadic and triadic situations.

Autism:

Autism is a neurocognitive disorder. Because the effects of autism can range from severe to mild, autism is considered as a spectrum disorder (Bowler, 2012). Autism is characterised by repetitive and stereotypical behavior apparent by 3 years of ages; deficits in verbal and nonverbal processes; deficits in emotional and social interaction.

Its genetic and neurocognitive aetiology is unknown. However different hypotheses have and continue to be discussed.

Autism is considered a complex multifactor disorder involving many genes.

 ISBN: 978-1-61208-250-9

Neuroimaging studies have demostrated that different brain areas are involved in autism. These studies reported aberrant brain connectivity and disruption of white matter tracts between temporal regions. They also shown abnormal activity in the exterior, the interior temporal and in the orbital frontal cortex (ventral part of the prefrontal cortex).

This data could provide the basis for concluding that in autism the more impaired cortical areas are those that are involved in complex cognition functions such as perception, interpersonal interaction and emotion, disorders which characterize autism.

Artificial environment such as robots have been used in the education of autistic children. They seem to be more effective than real environments.

Different projects exist.

The Aurora project study investigates the use of robots (Labo-1, Kaspar, Robota doll, for example) in game play. The aim is to create a tool based on an autonomous robot that convinces autistic children to engage in a process of interaction. The interactions are tested through the analysis of visual contact, joint attention, avoidance or fleeing, visual pursuit, and whether the child imitates the robot. Using a variety of modalities for interaction such as music, color and visual contact, a sensitive robot named Tito was employed in social interaction with autistic children. A very small fixed robot named Keepon captured and maintained visual contact with the child, drawing his attention and initiating some element of conversation.

 ISBN: 978-1-61208-250-9

Roboto used the form of an animated face (mouth, eyebrows, eyes) that, can cause behavior imitation from the part of the autistic child. Pleo is a dinosaur whic is used to encourage social interaction.

All these projects have reported dyadic interaction between the autistic child and the robot. They have shown that animate robots, humanoid or not, using different stimulations, encourage interaction in autistic children.

Even if quantitative metrics of social response for autism diagnosis including robots were developed, only one study has used a quantitative technique for analysing dyadic interaction. With the exception of Labo-1 in the Aurora project, and Rollball (Michaud's project) so far, only fixed robots have been utilized reducing the child’s spontaneity and self – expression in game play.

The aim of our studies was to analyse multimodal interactions of severely autistic children during free spontaneous game play with a mobile toy robot named «GIPY 1» both in dyadic child-robot situation and triadic child-robot-adult situation.

DYADIC INTERACTION

In the dyadic situation, we hypothesized that the autistic child will be in quasi-constant interaction with the robot.

 ISBN: 978-1-61208-250-9

The experimental set up was the following: The robot was manipulated by an operator via a wireless remote control using a joystick connected to a laptop. The robot could move forward, backward and turn on itself at low speed. These movements did not vary from child to child. The duration of the session was 5 minutes. The game play session began in exactly the same way for each child. When the child and tutor entered in the room, the tele-operated robot carried out three movements (move forward, move back, 360° swivel). As in real social interaction, the child and the robot altered their responses. If the child approached, the robot moved back and conversely. If the child moved away from the robot, i.e., ignored the robot, the robot followed the child in order to attract his/her attention. If the child remained motionless, the robot approached or turned around in order to focus the attention of the child. All the movements were standardized across the children.

4 children were observed during five minutes.

In this study, we analyzed the time of child-robot interaction (dyadic interaction). This time was defined as the duration between the onset time and the offset time of each child's behavior toward the robot. Four criteria were defined. These criteria were: 1) eye contact (looking at the robot), 2) touch (touching the robot without manipulating it), 3) manipulation (operating the robot), and 4) posture (changing corporal position toward the robot).

 ISBN: 978-1-61208-250-9

The duration of each criterion was calculated in seconds and was considered independently of the others. Concerning, for example, the characteristic "s/he looks at the immobile robot" ("eye contact") the onset time corresponded to the time when the child looked at the robot and the offset time to the moment when the child looked away from the robot. Accordingly, we calculated the duration of all the characteristics of each criterion. We summed up the duration corresponding to each criterion for each child. Only the total duration is presented in the results section.

Our results show that the children spent more than 79% of their time (156 seconds for the first, 289 seconds for the second, 269 seconds for the third and 241 seconds for the fourth child) playing with the robot.

As autism is a spectrum disorder where a large variation in abilities and interests among autistic children is apparent. At an individual level, the duration of "eye contact" is similar for all the children. However the analysis of the duration of "touching", "manipulating" and "posture" possibly reflects individual differences among the children. This analysis also showed how autistic children's behavioral interaction with the robot changes over a period of time. In other words, this suggests that a mobile toy robot could help autistic children to reduce repetitive and stereotypical behavior.

It seems that free game play could be a relevant ecological situation where an autistic child spontaneously interacts with the robot.

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-250-9

These findings tell us that an ecological situation, i.e., free game play, which is very closed to a real life situation encourages an autistic child to interact with the robot in a spontaneous manner. They also reveal that free, spontaneous game play with robots is possible with autistic children and could better facilitate the transfer of social and learnt abilities to real life.

What is important to demonstrate is whether and how autistic children can generalize learnt abilities during play with the robot to therapists and parents, i.e., proving that the robot could be used as a neural mediator tool for the enrichment of child-human interaction.

TRIADIC INTERACTION

This is what we tried to analyze in a third case study involving triadic child-robot-adult interaction.

We used the same criteria as above.

Based on the hypothsesis that cognitive multimodal interaction could be leed to the expression of an emotional state an additional fifth criterion was defined for the triadic interaction. This criterion is: positive emotion (display of enjoyment).

The above children were observed for 5 minutes of free game play.

 ISBN: 978-1-61208-250-9

The results show that the child spends half the time playing with the robot and half the time playing with the robot and the adult.

The duration time of “eye contact” and of “touch” is quite similar in both situations. However the duration time of “manipulation”, of “posture” and of “positive emotion” differ between the two situations. Positive emotion is more easily expressed when the child interacts with the adult and the robot than when the child interacts only with the robot. This difference reflective of the changes in autistic child behavoir with the robot could be used as a mediator for social and emotional interaction.

The robot, i.e., the artificial environment, seems mediate the interaction between autistic child and adult once the robot-child interaction is established.

In other words, cognitive and emotional multimodal interactions could also play a role in autism neurorehabilitation.

Robot neurorehabilitation might have high potential for improving the brain activity of child with autism. In this context, the robot could be considered as a « neural orthesis ».

In other words, artificial environment rendered possible through the use of mobile toy robots could lead to the neurorehabilitation of autistic children or atypically developing children.

Certainly, a robot could be considered as a orthesis (όρθεσις) used as a social stimuli mediator, with the ability to activate the same brain areas sensitive to humans in order to reduce the impairment of skills related to social and emotional information processing.

Thank you for your attention

 ISBN: 978-1-61208-250-9

What Should a Robot do for you? - Evaluating the Needs of the Elderly in the UK

Hagen Lehmann❖, Dag Syrdal❖, Kerstin Dautenhahn❖, Gert Jan Gelderblom★, Sandra Bedaf★ and Farshid Amirabdollahian❖

University of Hertfordshire❖
h.lehmann@herts.ac.uk
d.s.syrdal@herts.ac.uk
k.dautenhahn@herts.ac.uk
f.amirabdollahian@herts.ac.uk

Zuyd University★
gertjan.gelderblom@zuyd.nl
sandra.bedaf@zuyd.nl

***Abstract*—The increasing interest in the use of robotic assistive technologies in elderly care for the UK makes it necessary for roboticists to evaluate the needs, problems and demands of possible end-users of such technologies. Users of these technologies can be divided into three groups: informal caregivers (family members and friends), formal caregivers (medical staff, social workers, home-help), and the elderly themselves. In this paper we present the results of a series of focus groups conducted between March and May 2012. We used the metaplan method to evaluate the opinions and needs of each of the three different potential user groups mentioned above. In these discussions we extracted a variety of problem dimensions and their interconnections in order to understand in which parts of everyday life assistive technology could help, and is needed the most.**

***Keywords*-elderly care; evaluation of needs; robotic assistive technology**

I. INTRODUCTION

The general aim of the work presented in this article is to develop scenarios and socially acceptable behaviours for a service robot that can be used for home assistance for elderly people, to facilitate longer independence for them at their own homes. In order to evaluate what a robot in such a setting should or should not do and how it should behave towards the users, we involved different potential user groups right from the beginning of the project. Our user-centred approach will help to make sure that the results will realistically reflect the everyday experience and needs of the potential end-users of assistive robotic technology.

This work was undertaken as part of the EU FP7 Project ACCOMPANY (Acceptable robotiCs COMPanions for AgeiNg Years) [1]. The goal of this project is to create a system in which a robot companion is integrated within a smart-home environment in order to facilitate independence at home.

We will start our paper by giving an overview of the motivation for this research, based on the demographic changes in the western population and a description of the ACCOMPANY project. This will be followed by a description of the metaplan method and the experimental structure we used. At the end of the paper we will present our results and discuss their implications for the design process of robot home companions.

II. MOTIVATON

The demographic shift presented by falling birth-rates and increasing life-spans is leading to an ageing population worldwide [2]. In the United Kingdom it is expected that by 2050 one person in four will be above the age of 65, and one person in twenty above the age of 85 [3]. These changes are presenting challenges to the way that geriatric care is provided.

Many researchers are suggesting that a serious consideration in how such care is provided and organised is needed [4]. This will necessarily involve further adoption of technological solutions, tele-care, along with smart-home sensor environments which are already proving to be an effective means for people to maintain independent living for longer [5]. Robotics technology is particularly attractive as it allows for physical interactions and practical assistance within the home environment. Roy et al. [6] argue that, from a technological perspective, the falling cost of sensor technologies and computing power increasingly bring personal robotics in eldercare into the realm of feasibility and also highlights a series of applications that a personal service robot may have.

Within the ACCOMPANY project we have defined 3 groups of users. (1) Professional caregivers, such as nurses, social workers, home helpers and care assistants, (2) informal caregivers, such as relatives, spouses, neighbours and friends, and (3) the elderly themselves. Recognising that these three groups may represent three distinct experiences of the issues facing the elderly, we realise that there is a need to consider all three groups. Previously, researchers in robotics have examined these perspectives from a more general point of view. Tsui and Yanco [7] examined applications and attitudes towards robots in healthcare amongst professionals, and found that this user group primarily saw robots as work-tools that could be used by the carer to aid in the performance in specific tasks, like lifting tools, deliveries and for information sharing. When examining the views of the elderly, Cesta et al. [8] found that, while their preferred usage-scenarios addressed the realities of their everyday life, it was challenging to relate them to the realities of the available robotic platform, reducing direct applicability of these to the developers of an assistive robotic system. Harmo [9] also recognises the distinction between carers and the elderly, and their findings echo those of Tsui and Yanco in terms of how professional carers viewed the role of an assistive robot. Harmo also highlights the importance of directly relating findings from potential users to possible technical solutions.

III. APPROACH

We are aiming to examine user needs and translating these into specific usage-scenarios using an iterative process in which the findings arising from consultation with potential users are tied directly to the technical development within the ACCOMPANY project in a manner similar to that which we have done previously with other types of user-groups [10]. This iterative approach allows for meaningful user involvement at all stages in the development of the system.

In order to evaluate the perspectives of the three user groups mentioned above, we organised a first set of focus groups using the metaplan-method [11]. The metaplan-method aims at defining different problem dimensions in a moderated discussion amongst group members. The idea is to use the creativity and interaction dynamics of the group members to extract ideas from the group, ideas that single members might not have been aware before the brainstorming. To create this kind of group dynamics the minimal size of the group should be 4 or more.

We used a three-step approach. We started by having each group member write down the issues and specific problems they think are important independently on post-it notes. Second, all these notes were put on a white board and then organised in a discussion by the group members into problem clusters, which were then defined as different problem dimensions. The last step was to rate these problem dimensions, and discuss possible connections between them. The professional caregiver group consisted of 4 women, in the group of the informal caregivers were 5 women and in the elderly people group were 2 women and 3 men. The average age of the elderly participants was 76.2 years, ranging from 70 to 83 years. An actual robot was not mentioned in the focus groups, instead the more general question “What everyday problems threaten independent living for elderly people?” was asked.

IV. RESULTS

A. Problem dimensions

In a first step different problem dimensions were identified in each of the user panels. Most of these general problem spaces were similar between the groups. Yet some of them differed, depending on the priorities and perspectives of the groups.

B. Professional caregivers

In order to fully explore the insights the professional caregivers have into the issue, we used a two-step approach with this group. We conducted a metaplan focus group at the University and then used the findings of this user panel to inform a second discussion with employees directly at a local care facility. During focus group discussion the formal caregivers agreed on 7 main problem dimensions - Environment, Physical Health, Mental Health, Communication, Society/Family, Personal Traits and Self-confidence.

Each of these dimensions was discussed in detail and for each a list of specific problems and subcategories of problems was given. The environment dimension was specified as problems with modern living, finances, access to services and lack of family and friends. As specific problems for physical health, the nutritional status, complicated medical treatment, incontinence, not being able to cook anymore, inability to manage house, loss of senses (mainly sight, touch, hearing), fear of falling, inability to dress, immobility and loss of strength, as well as dexterity were given. Mental health problems were named as inability to retain instructions/guidance, Alzheimer/dementia, loss and bereavement resulting in depression and anxiety, forgetting to eat and memory loss. According to this group, examples of problems with communication are the inability to understand the jargon used in IT and the Internet, as well as by doctors and nurses. Examples of problems with family and friends, which could challenge everyday independent living, were defined as isolation, loneliness, the fact that the family cannot cope with the situation, a lack of support, and problems with managing the personal financial situation. Personal traits that could be a problem for independent living were specified as the inability to recognise the need for help, not to want help in general, low frustration tolerance ("I can't do it, so I won't anymore") and the opinion that one doesn't want to be a burden.

At the end of the discussion we asked the group members to specify the interactions between each of the problem dimensions. In the case of the professional caregivers the resulting structure was very interesting. We found that they defined almost all dimensions as being in interaction with each other, and that the lack of confidence is the central issue that threatens independent living of elderly people in everyday life most (Figure 2).

C. Informal caregivers

For the informal caregivers we held the focus group at the University. This group contained people who were caring for one or more relatives at their home. The focus group uncovered 9 primary problem dimensions - Emotional Situation, Physical Decline, Cognitive Decline, Economic Situation, Infrastructure, Reluctance to use Technology, Security, Family and Society/Public Attitude/Policy.

As in the professional caregiver group, each of these dimensions was illustrated by a list of specific problems. The problems defining the emotional situation were specified as fear of the future, fear of family interference, negative moods (e.g. anger, irritation, etc.), resistance to appropriate help from others, resistance to helpful adaptations, a lack of subtleness and/or sensitivity on the caregivers' side, frustration not be able to do things independently, having to wait until someone can do things for you, and the perception that time passes slowly when one cannot leave the house. The physical decline dimension was defined by mobility problems, loss of senses (mainly sight, hearing, touch), incontinence, loss of dexterity, poor arm strength and flexibility, as well as balance issues. Problems constituting the cognitive decline were given as memory loss, inflexibility of ideas, dementia and suspiciousness increasing with the age. Problems in the economic situation were defined as caused by an insufficient income to pay the type of help wanted, and income issues in general. According to this group, they would result in problems with shopping and eventually eating. Infrastructure was another problem dimension, specified as the lack of transport and accessibility of the public space, poor and unsuitable housing conditions, as well as limited access to hospitals.

The reluctance to use technology was specified as a dislike of "modern gadgets", unfamiliarity with the proper use of new technology, and timidity about using technology. According to the informal carers, security problems have two aspects.

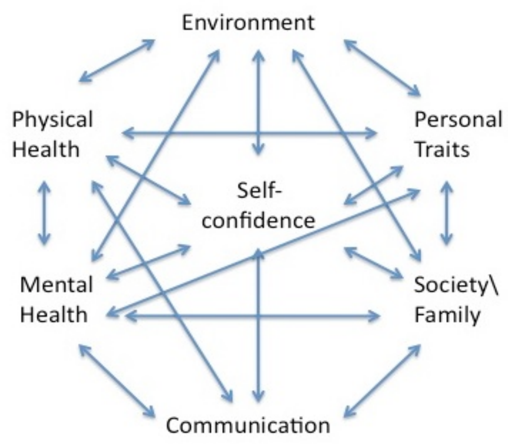


Figure 2: Problem dimension space for "Professional Caregivers"

One is the individual feeling of being insecure about moving around in the house (e.g. fear of falling), and the other is the fear of intruders. Issues with family and friends could be that they live too far away, that the caregiver wants to organise and control the person, that the person hides things from the caregiver for fear of being put into a home, and that the caregiver does not want the person to be aware of the fact it is hard to care for her. The last two problems can be summarised as an atmosphere of secrecy between family members. These problems are closely related to what was described as the society/attitudes/policy dimension. According to the unofficial caregivers, its main factors are isolation, the feeling of loneliness, the attitude of other age groups towards older people, and the need to find various professionals (e.g. gardener, hairdresser, etc.) to come to the home.

The interaction structure of the problem dimensions differed significantly from the one given by the professional caregivers. The informal caregivers did not connect all the dimensions with each other and also did not assign a central role to any of the dimensions. The emphasis for them was on the emotional situation of the elderly person, which, according to them, was influenced by, and influenced most of the other problem dimensions (Figure 3).

C. Elderly people

The focus group for the elderly was arranged at the University as well. The participants were on average 75 years old and not younger than 65. For this group we found 12 different main problem dimensions - Physical Health, Emotional Health, Family/Friends, Mobility, Security, New Technologies, Communication, Services, Transport, Public Finance Control, Personal Finance and Lack of Trust in General.

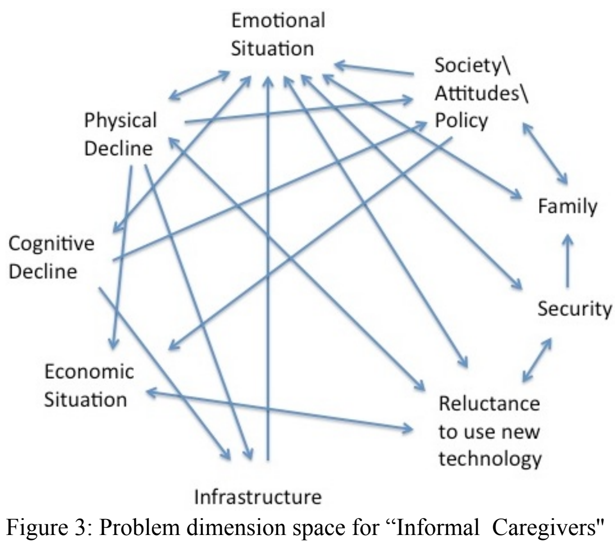


Figure 3: Problem dimension space for "Informal Caregivers"

The deterioration of physical health was the problem discussed by the elderly in most detail. A lot of examples were given to specify the general issue of becoming old and less able to do things: the loss of sight (limiting reading for any length of time as well as other similar tasks), problems with knees, hips and other joints (making it difficult to get out of bed, impairing seating positions, making it difficult to kneel down, to climb stairs, to carry heavy things, like shopping bags around the house), difficulties of getting up from low lavatory seats or chairs without arms, the fact that arthritis makes gardening and housekeeping difficult and that cooking becomes problematic. These problems are closely related to the dimension of services and carers. The elderly find it problematic to hire someone to help with general housekeeping and gardening. Getting this sort of help usually implies to let someone unfamiliar into the house, and thus creates a feeling of insecurity for the elderly.

The problems for emotional health were defined as the worry about how to cope with the potential necessity of downsizing or going into a care home, the fear of how to cope with the death of the partner and the reliance on home care. As the main problem with family and friends, they indicated the absence of their proximity. Mobility was specified as another problem dimension: difficulties with stairs, impossibility to drive a car and the related restricted travel possibilities in small villages, the loss of general physical mobility. Issues with public transport were identified as closely related to these problems. The lack of an efficient public transport system, and the general limitations of public transport were named as main issues. The problem of security was described as the general concern of the criminal element. For the elderly new technologies represent another aspect of everyday life that causes problems with independent living. The reliance on computers in the public domain, and the limited knowledge they have of these devices, causes the elderly to worry. As an example, they pointed out the difficulty of scheduling a GP appointment online. Problems with communication with official entities due to a incomprehension of specialised language was named as another issue potentially limiting independent living. For example, the use of pro-forma letters without a specific point was mentioned. Another problematic aspect pointed out by the elderly was the financial situation. This problem dimension was defined as two-sided. On the personal finance side, problems were considered to arise due to changes of pensions and the resulting insufficient funding. On the public finance side, the two big problems mentioned were the feeling of not being in control of ones own finances and the need for a better health system. The last important point mentioned during the elderly group discussion was their perception that it has become increasingly difficult to trust someone, due to developments in the society.

The specific interaction structure given by the elderly between the problem dimensions was more similar to the one given by the informal caregivers than to the one given by the professional caregivers. Despite the fact that they did not connect all dimensions with each other, and did not assign a central role to one of the dimensions, they also did not emphasize one of the problems (Figure 4).

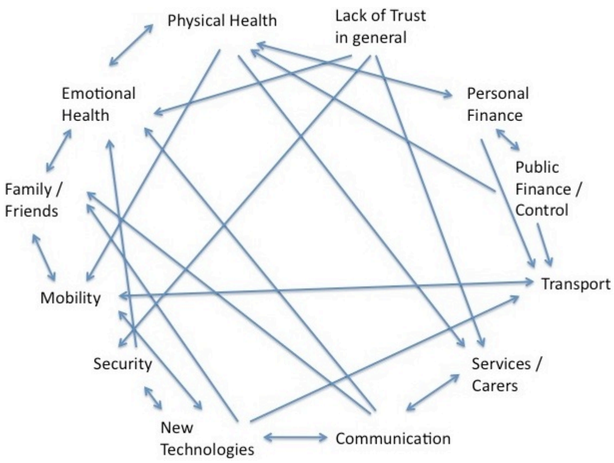


Figure 4: Problem dimension space for the “elderly"

D. Similarities and differences

There are five main problem complexes, which were mentioned by the three user groups. These are the decline of [1] physical health, [2] the economic situation, [3] family and friends and [4] transport, mobility and infrastructure.

Both the informal and the professional caregivers additionally mentioned the cognitive decline or mental health, and attitudes in the society in general. For the informal caregivers, as well as for the elderly, the emotional situation, security issues and the need to find service persons like hairdressers or gardeners to come into the home are substantial problems. The professional caregivers and the elderly mentioned the inability or lack of communication as one of the most important factors threatening independent living in everyday life. Some of the problem dimensions were only mentioned by one of the user groups. The professionals mentioned personal traits as a problem, and emphasised on self-confidence as a central issue. The informal caregivers singled out the reluctance of the elderly to use modern technology as a central issue. The elderly focussed on the lack of trust in contemporary society, a general fear of the criminal element and the use of new technologies in society.

In general it can be said that physical health and the problems related to limited physical mobility are the biggest and most important problematic dimensions. Issues in house keeping and personal care related to arthritis, loss of dexterity, joint pain and the loss of senses are the most prominent specific problems mentioned. This is closely related to the absence of friends or family, which were also mentioned by the three user groups. This is not very surprising, but strengthens the point that special robotic home companion technology can be a great help in assuring independent living of the elderly.

The most interesting differences are that for the group of elderly people there is an emphasis on trust towards unknown people and uncertainty about personal safety with regard to the criminal element in society. It seems that older people feel more threatened by their environment. Another point is that the elderly are seeing the existence and the widespread use of new technologies as one of the major problems, whereas the informal caregivers see the reluctance on the side of the elderly to use new technologies as the problem. These findings are consistent with several results of studies and surveys concerned with the “grey” digital divide [12, 13]. The formal carers on the other hand put a lack of self-confidence elderly people have about their abilities in general in the centre of their thoughts. In their opinion all other dimensions are influenced by this factor.

In general most of the specific problems mentioned by all three groups are similar and differ only by in their categorisation. For example, the financial situation was categorised as a society issue by the professional caregivers, but for the informal caregivers and for the elderly it was an independent problem dimension. These categorisation differences reflect the importance given to the problem by the corresponding user group.

V. DISCUSSION

The problems immediately relevant to the ACCOMPANY project are the issues in house keeping, arising from limited mobility and decreased dexterity due to deteriorating health, as well as security, communication, the reluctance to use new technologies and psychological problems arising from being lonely.

Current service robots are developed towards achieving functions such as safely navigating around the home and helping with some tasks in the house, reminding users to take medicines and to eat regularly, helping with some kitchen work and serving, and potentially helping with mobility issues, e.g. opening doors. Note, in particular those activities that involve the robot carrying out complex manipulations of objects are still highly challenging from a technological perspective. These abilities, if implemented efficiently and for safe operation, could potentially aid everyday life and thus improve the life quality of elderly people.

Among different robotic platforms, for example the Care-O-bot® 3 represents a potential possibility to be used as a companion in a domestic environment [14]. It can draw on a sophisticated set of sensors, enabling it to detect people, detect and recognise some objects in its environment, as well as safely navigate in environments where it has to co-exist with humans.

At the University of Hertfordshire Robot House [15], these capabilities could be joined within an ecologically valid testing environment, which allows for the detection of activities through different sensors situated in the house itself. Together, these features would potentially allow for a wide set of functionalities and usage-scenarios.

The integrated nature of the robot house sensor networks, and their ability to communicate with the robot itself, implies that it can collate a substantial amount of information and to utilise this information for the benefit of different user groups. In other words, it can alleviate a considerable amount of stress that the informal carers expressed regarding not knowing how much help the assisted elderly actually needs, and potentially address some of the security issues [6] mentioned by the informal caregivers. Of course there are serious ethical concerns associated with this type of information sharing with third parties [16], which we aim to address in subsequent user studies.

A robot companion can also help with problems of loneliness, especially when equipped with individual features, which allow the user to identify themselves with their robot. One possibility to achieve this would be to enable robot customisation. Results from studies with the Roomba robotic vacuum cleaner illustrate that people start to describe the robot in aesthetic and social terms [17]. Additional the use of tele-presence robots has shown that people intuitively start to individualise the robot they are interacting with. Ethical issues arising from a potential emotional attachment to a machine will also have to be examined and evaluated carefully in structured user studies.

Ultimately, a combination of smart sensors and a robot companion would be helpful in the context of care. Our evaluation has shown that we need to take three different perspectives into consideration and not only focus on the needs of the elderly, but to address the needs of professional and informal caregivers, as they also play an important role in the integration of robotic assistive technology into elderly care. The differences in the problem dimensions we have found in our evaluation illustrate this very clearly. Any new technological development needs to be accepted by the key stake-holders involved. This is a key element, in addition to e,g. cost effectiveness and service delivery models, for bringing these systems into the real world.

Since both the informal caregivers and the elderly mentioned the reluctance or inability to use new technologies as a significant problem, it seems that improving the elderly people’s acceptance of robotic companions is going to be one of the immediate issues, if an integration of robot companions into the elderly care process is going to be successful. To address this particular problem, we plan to involve potential users in extensive Human-Robot Interaction studies in the robot house at the University of Hertfordshire as part of the later stages of the of the ACCOMPANY project.

VI. CONCLUSION

The results reported in this article represent the first step of the design cycle, whereby the described user panels were conducted with different stakeholders, without any information given about possible concrete robots to be used in this work, so as to avoid a bias of their views and attitudes due to the technology. Further rounds of user panels will follow to refine the tasks and scenarios throughout the duration of the project.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the [European Union's] Seventh Framework Programme (FP7/2007-2013) under grant agreement n°[287624].

REFERENCES

[1] ACCOMPANY. (16.01.2013) ACCOMPANY website, url: http://accompanyproject.eu

[2] United Nations. (2002) World Population Ageing 1950-2050 .United Nations, New York.

[3] Cracknell, R. (2010) The Ageing Populaton. "Key Issues for the New Parliament 2010". *House of Commons Library Research.* pp. 44 - 46.

[4] Kovner, C.T., Mezey, M. & Harringdon C. (2002) Who cares for older adults? Workforce implications of an ageing society. *Health Affairs*. 21(5), pp. 78-89

[5] Magnusson, L., Hanson, E. & Borg M. (2004) A literature review study of Information and Communication Technology as a support for frail older people living at home and their family carers. *Technology and Disability*. 16(4), pp. 223-235

[6] Roy, N., Baltus, G., Fox, D., Gemperle, F., Jennifer Goetz, J., Hirsch, T., Margaritis, D., Montemerlo, M., Pineau, J., Schulte, J. & Thrun, S. (2000) Towards Personal Service Robots for the Elderly. *2000 Workshop on Interactive Robots and Entertainment, April 30, 2000, Pittsburgh Pennsylvania,* pp. 184

[7] Tsui, K. & Yanco, H. (2007) Assistive, Surgical, and Rehabilitation Robots from the Perspective of Medical and Healthcare Professionals. *Proceedings of the AAAI Workshop on Human Implications of Human-Robot Interaction, at AAAI-07, July 22-26, 2007, Vancouver, Canada.*pp 34-39.

[8] Cesta, A., Cortellessa, G., Giuliani, M.V., Pecora, F., Scopelliti, M. & Tiberio, L. (2007) Psychological implications of domestic assistive technology for the elderly. *PsychNology Journal.* 5(3), pp. 229-252

[9] Harmo, P., Taipalus, T., Knuuttila, J., Vallet, J. & Halme, A. (2005). Needs and Solutions - Home Automation and Service Robots for the Elderly and Disabled. *IEEE/RSJ International Conference on Intelligent Robots and Systems 2005, 2-6 August 2005,Helsinki, Finland,* pp. 3201-3206

[10] Robins, B., Ferrari, E., Dautenhahn, K., Kronreif, G., Prazak-Aram, B., Gelderblom, G.J., Bernd, T., Caprino, F., Laudanna, E. & Marti, P. (2010) Human-centred design methods: Developing scenarios for robot assisted play informed by user panels and field trials. *Int. J. Hum.-Comput. Stud.* 68 (12), pp. 873-898

[11] Schnelle, E., Thiersch, M. (1979) The Metaplan-Method: Communication Tools for Planning and Learning Groups. *Quickborn*, *Hamburg*

[12] Kiel, J.M. (2005) The digital divide: Internet and e-mail use by the elderly. *Informatics for Health and Social Care*. 30 (1), pp. 19-23

[13] Morris, A., Goodman, J., Brading, H. (2007) Internet use and non-use: views of older users. *Univ Access Inf Soc*. 6 (1), pp. 43-57

[14] Parlitz, C., Hägele, M., Klein, P., Seifert, J. & Kerstin Dautenhahn, K. (2008) Care-o-bot 3 - rationale for human-robot interaction design, *Proc. of 39th International Symposium on Robotics (ISR), Seoul, Korea,* pp. 275-280

[15] Koay, K.L., Syrdal, D.S., Walters M.L. & Dautenhahn, K. (2009) Five Weeks in the Robot House - Exploratory Human-Robot Interaction Trials in a Domestic Setting. *Proc. Of Second International Conference on Advances in Computer-Human Interaction, ACHI 2009 February 1-7, 2009, Cancun, Mexico,* pp. 219-226

[16] Anderson, M. & Anderson, S.L. (2008) ETHEL: Toward a Principled Ethical Eldercare Robot, *Proc. Robotic Helpers: User Interaction, Interfaces and Companions in Assistive and Therapy Robotics – Workshop at the Third ACM/IEEE Human-Robot Interaction Conference (HRI08), March 12-15,2008, Amsterdam, the Netherlands,* pp 33-39.

[17] Forlizzi, J. (2007) How robotic products become social products: an ethnographic study of cleaning in the home, *Proc. of the Second ACM/IEEE international conference on Human-robot interaction (HRI07), March 9-11, 2007, Washington DC*, pp. 129-136.

Investigating Child-Robot Tactile Interactions: *A taxonomical classification of tactile behaviour of children with autism towards a humanoid robot.*

Ben Robins, Farshid Amirabdollahian, Kerstin Dautenhahn
School of Computer Science
University of Hertfordshire
Hatfield, U.K
{b.robins, k.dautenhahn, f.amirabdollahian2}@herts.ac.uk

Abstract—**The work presented in this paper is part of our investigation in the ROBOSKIN project. One key research activity in the project was to explore tactile interactions of children with autism with the humanoid robot KASPAR in order to develop methods and mechanism to support robot-assisted therapy for children with autism. This article presents a detailed taxonomical classification of tactile interactions of 14 children with autism with the humanoid robot KASPAR. Our quantitative analysis confirms results from the literature highlighting the great variety of autistic children's interaction capabilities.**

Keywords-assistive technology; human-robot interaction; autism therapy; robot assisted therapy.

I. Introduction

Physical touch is one of the most basic, but at the same time very important forms of communication. Tactile sensing can help to provide awareness of one's own self and each other. On the playground, touch and physical contact are used by children to give and receive support and encouragement, to build trust, to communicate and to develop their social relationship. In therapy, the tactile sense can be used individually to increase self knowledge, body image, to achieve sense of stability, and build confidence. Touch of another person, when it happened, is seen also as a way of breaking through isolation [1, 2].

However, for people with autism, in addition to their inabilities to relate to other people, show little use of eye contact, and have difficulty in verbal and non-verbal communication, impairments in tactile interaction prevent them even further from social interaction with other people.

Some people with autism might be hyposensitive and seem not to feel pain. They may not sense their touch of other people or objects appropriately, which could lead them to unintentionally hurt other people, or break objects. Other people with autism might have a hyper-tactility condition which is very common and results in overwhelming sensation. As touch can be excruciating to people with this condition it leads to fear of being touched. This fear could be so great, that it may cause a panic attack [3, 4].

On the other hand, tactile interaction (if tolerated) might be an important means of communication for children with autism, as some do not have verbal skills and others use their verbal skills inadequately. Caldwell [5] suggests that problems with verbal skills and eye gaze in children with autism create the need for touch to replace these detrimental ways of communicating.

We argue that a 'tactile' robot can be used as a 'buffer' that mediates between a person with autism and another person, by providing indirect rather than direct human-human contact, until such time that the person builds enough strength and confidence to tolerate direct human contact.

A robot with tactile applications could allow a person with autism to feel safe and build their confidence in tactile interaction where they can explore touch in a playful way that could be completely under their control. Also, while inappropriate tactile interaction with another person will automatically lead to negative feedback (e.g., when hitting another person), interaction with a robot can provide a non-judgemental environment where the child can safely explore tactile interaction, in a long-term process that involves reflection and feedback given to the child about his or her actions.

Several observational systems and taxonomies can be found in the literature that were created to help clinicians and researchers making inferences about play, based on the observation of behaviours, e.g., Knox play scale [6], or Bundy's Test of Playfulness [7]. Taxonomies of children's behaviour during play provide criteria to guide observation of behaviours in narrower categories that may be easier to observe, describe and explain.

However, play behaviours at their core are thought to involve experiential characteristics that are typically difficult to observe directly, such as intrinsic motivation, enjoyment and active engagement, and it is important to remember that taxonomies often rely on observable behaviours which often cannot capture fully the person's subjective experience [8]. This becomes an even more important factor to consider when working with a population for whom communication skills are one of the main areas of impairment (such as the case of children with autism - many have limited or no language skills at all) and it is very difficult, if not impossible, to interview the person about their experience.

In addition to impaired communication, atypical sensory processing, motor difficulties, and cognitive impairment are other very common characteristics of autism. As children with autism may manifest these symptoms to varying degrees, this results in an extremely heterogeneous population [8], which in turn, makes the task of developing a taxonomy and classification of typical tactile behaviour of children with autism very difficult. Although children with

 ISBN: 978-1-61208-250-9

autism share the same core difficulties, each child displays these in an individual way [9]. A series of 14 interaction sessions that were conducted in a recent study with autistic children and the robot KASPAR were used as a basis for our taxonomical classification. The following sections describe the robotic platform used and the trials set up and procedures and continue with the taxonomical classification of tactile interaction followed by discussion and future plans.

II. THE ROBOTIC PLATFORM - KASPAR

KASPAR is a child-sized minimally expressive robot which acts as a platform for HRI studies, using mainly bodily expressions (movements of the hand, arms and facial expressions) and gestures to interact with a human (Figure 1). The robot has a static body (torso, legs and hands were taken from a child-sized commercially available mannequin doll) with an 8 DOF head and two 3 DOF arms.

The face is made from a silicon rubber mask that covers an aluminum frame. It has 2 DOF eyes fitted with video cameras; eye lids that can open and shut and a mouth capable of opening and smiling. These features enable KASPAR to show minimally expressive emotional states such as *happiness*, *neutral*, *sadness* and *surprise* (Figure ***2***). It has several pre-programmed behaviours that include various facial expressions, hand waving and drumming on a toy tambourine that is placed on its legs. KASPAR's movements are either controlled remotely through a remote control device, or it can operate autonomously.

For a complete description of KASPAR's design rationale, hardware, and application examples see [10].

III. TRAILS SET-UP AND PROCEDURES

The study, which involved 14 children with autism, was designed to provide essential observational data on children's behaviour during child-robot interaction including spontaneous tactile interaction.

The trials took place in a special needs school for children with moderate learning difficulties in the UK. With the objective to provide a reassuring environment, the trials were designed to allow the children to have free and unconstrained interaction with the robot and with the present adult (i.e., teacher, experimenter) should they wish to. The trials were conducted in a familiar room often used by the children for various activities. Before the trials, the humanoid robot was placed on a table, connected to a laptop. The investigator was seated next to the table. The robot was operated remotely via a wireless remote control (a specially programmed keypad), either by the investigator or by the child (depending on the child's ability). The children were brought to the room by their carer and the trials stopped

Figure 1. The robotic platform KASPAR.

Figure 2. Some of KASPAR's facial expressions and expressive gestures.

when the child indicated that they wanted to leave the room or if they became bored. Two stationary video cameras were used to record the trials.

IV. TAXONOMICAL CLASSIFICATION OF TACTILE INTERACTION

All interaction sessions of the children and the robot were video recorded from different angles. 14 sessions (one for each child) were used in building the taxonomical classification reported here. The resulting videos were analysed in a coding process that included watching the videos and manually identifying touch types, locations, durations and estimated pressure by observing video sequences and coding start, end or onset of events. This entailed complete coding of all 14 videos using the Observer software (Noldus). Table 1 shows the tactile features that were observed during the interaction. In addition, the duration of these features were detected and calculated.

A. Preliminary visual inspection of the coded events

In order to provide a more detailed classification and taxonomical breakdown, events coded in these videos were further analysed. Figure ***3*** shows a visual representation of the events detected in 4 of these videos. As can be seen, these four cases present a very different result in terms of type, duration, and frequency of the tactile events observed.

In order to gain a fuller picture of these variations, number and duration of these touch events were plotted versus one-another (Figure 4 & Figure 5). As different participants interacted with the robot for different lengths of time, the number of tactile events are normalized based on the durations. Thus a further plot is included presenting the rate of occurrence for these tactile events (Figure 6).

TABLE 1
TACTILE FEATURES: TOUCH TYPES, PRESSURES AND LOCATIONS

Touch type
Grasp
Touch
Stroke
Poke
Pinch

Touch Pressure
Light
Medium
Tight

Touch Location
Left arm
Left leg
Right arm
Right leg
Both arms
Both legs
Head
Torso

 ISBN: 978-1-61208-250-9

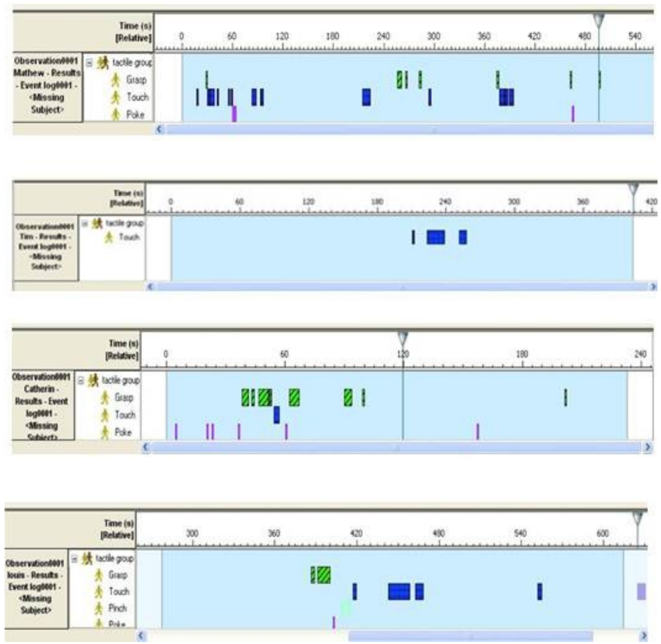


Figure 3. Four examples of tactile events detected.

In order to further investigate the extent of these observed differences, coded events were analysed statistically.

A. Statistical analysis of the coded events

Comparing the number of touch events as shown in Figure 3 can be misleading as longer sessions can include a larger number of tactile events. The occurrence rate of tactile events is a compound variable that consists of the number of specific tactile events and the total duration of each session thus allowing for comparison based on rate (number/duration). This parameter was analysed statistically in order to highlight differences in between touch types and individual participants.

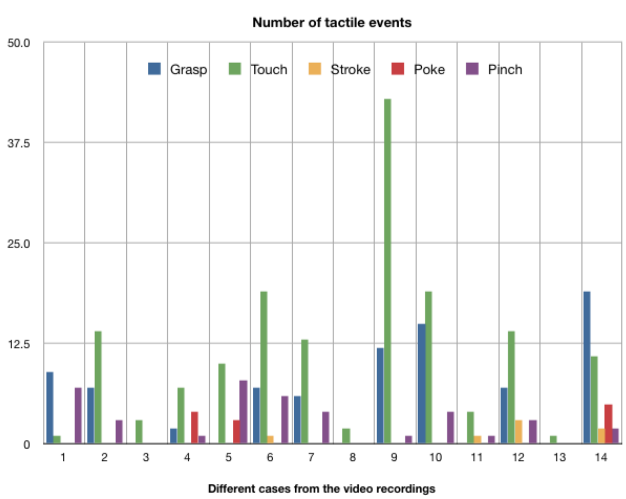


Figure 4. Comparison between number of tactile events detected for different cases.

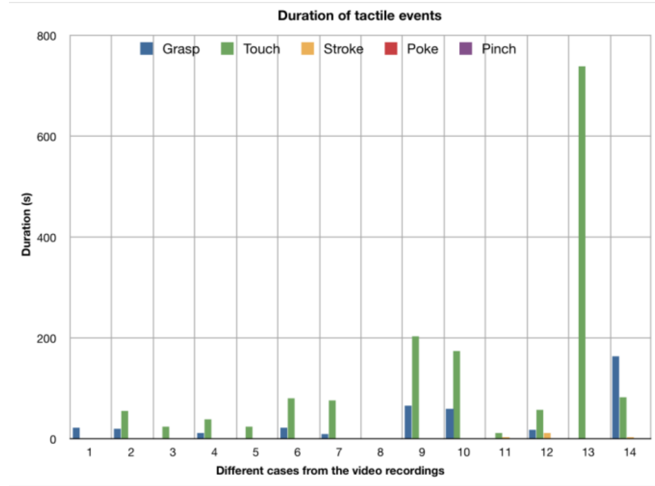


Figure 5. Comparison between the duration of the tactile events detected from different cases. Please note that poke and pinch were point events and did not have a duration attribute.

At a first glance, the boxplot presented in Figure 7 shows the extent of these observed variations as presented by the 'rate per minute' variable. As can be seen, the grasp and touch events present higher rates of touch and also higher variability while pinch, stroke and poke present lower rates, and less variability.

1) General linear model (two-way ANOVA)

In order to identify the extent of inter-event variations, as well as variations between each touch type, a two-way ANOVA model was constructed using the PASW statistical analysis package. This package uses the general linear model for multivariate analysis including the two-way ANOVA. Model variable was the 'Rate per minute', while factors were 'touch type' and 'participant'. This allows for identifying differences between different touch types and

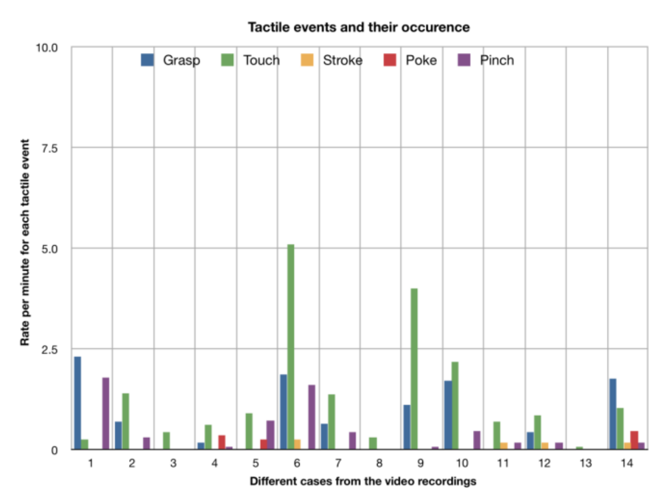


Figure 6. Occurrence rate per minute for each tactile event for different cases. These include normalised number of events over the duration of each session in minute.

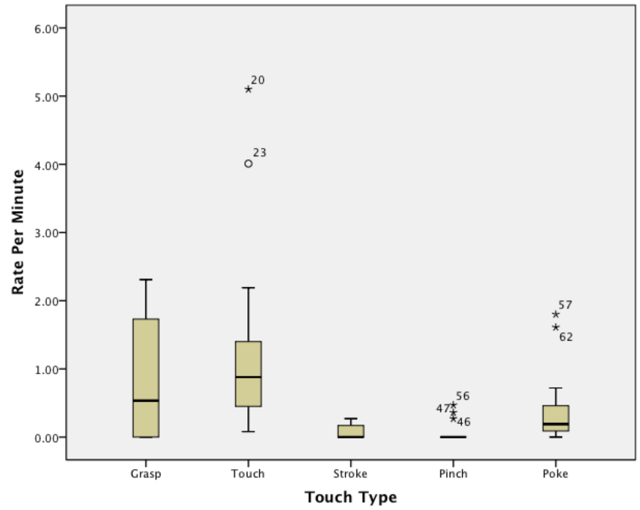


Figure 7. Boxplot comparing rate per minute for different tactile events.

also how participants performed under each type of tactile event.

Table 2 highlights that in addition to significant variations between the touch events identified, there were significant variations between different participants and their rate and type of touch.

Table 3 presents the parameter estimates and their influences on the general linear model. It shows that different parameters (intercept, participant 6 and touch type 2) had most influenced the linear model. Such a model is a type of linear best-fit line for the data points and here the model tries to fit a line to those data points presented by touch type and for different participants. Significant and influential parameters are shaded rows while close to significant values are shown by shading the cell only.

It is interesting to note that participant 6 and touch type 2 are presented as the most dominant features seen in Figure ***5***, while participants 8 and 13 present the least visible features in this figure, highlighted by the negative slope value (column B in Table 3).

TABLE 2
TOUCH TYPE AND PARTICIPANT VARIATIONS SHOWN BY A GENERAL LINEAR MODEL

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	53.819	1	.000
Participant	39.104	13	.000
TouchType	44.198	4	.000

Dependent Variable: Rate Per Minute
Model: (Intercept), Participant, TouchType

TABLE 3
PARAMETER ESTIMATION AND IDENTIFICATION OF SIGNIFICANT PARAMETERS

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.621	.3144	.004	1.237	3.897	1	.048
[Subject=1]	.142	.3921	-.626	.910	.131	1	.717
[Subject=2]	-.252	.3921	-1.020	.516	.413	1	.520
[Subject=3]	-.642	.3921	-1.410	.126	2.681	1	.102
[Subject=4]	-.482	.3921	-1.250	.286	1.511	1	.219
[Subject=5]	-.354	.3921	-1.122	.414	.815	1	.367
[Subject=6]	1.040	.3921	.272	1.808	7.036	1	.008
[Subject=7]	-.240	.3921	-1.008	.528	.375	1	.540
[Subject=8]	-.668	.3921	-1.436	.100	2.903	1	.088
[Subject=9]	.312	.3921	-.456	1.080	.633	1	.426
[Subject=10]	.144	.3921	-.624	.912	.135	1	.713
[Subject=11]	-.524	.3921	-1.292	.244	1.786	1	.181
[Subject=12]	-.398	.3921	-1.166	.370	1.030	1	.310
[Subject=13]	-.716	.3921	-1.484	.052	3.335	1	.068
[Subject=14]	0[a]	.	.	.	.	.	.
[TouchT=1]	.337	.2343	-.122	.796	2.070	1	.150
[TouchT=2]	.947	.2343	.488	1.406	16.340	1	.000
[TouchT=3]	-.374	.2343	-.833	.086	2.542	1	.111
[TouchT=4]	-.354	.2343	-.813	.106	2.277	1	.131
[TouchT=5]	0[a]	.	.	.	.	.	.
(Scale)	.384[b]	.0650	.276	.535			

Dependent Variable: Rate Per Minute
Model: (Intercept), Participant, TouchType
a. Set to zero because this parameter is redundant.
b. Maximum likelihood estimate.

2) *General linear model for tactile location and touch intensity*

We further extended the analysis by considering variations of touch across different body parts of the robot shown by TABLE 1. This is a more complex statistical model, which would allow investigating if there were similarities for touch event rates for a specific body part and specific touch intensities. Such a model requires a more detailed coding and calculation of rates per body location, as well as data structuring to allow for a two-way ANOVA, done under the PASW general linear model. The model constructed here uses the 'rate per minute' as its variable while factorising 'participant', and 'body part; intensity'. The last parameter identifies the body part touched and observed intensity of the touch type. The box plot presented by Figure ***8*** highlights these results.

The general linear model once more identified significant differences in rate per minute between participants and 'touch event; intensity', which is evident in the above boxplot (Table 4). These differences are highlighted by Figure 9 and TABLE 5.

Further results are provided by the parameter estimates (TABLE 5). This table shows that when modeling based on observed variations, participant 6 still presents a strong positive influence while participant 8 continues to provide a negative contribution towards a common fitted line. Other participants, participant 9 and 10, also provide significant contributions towards the trend. Touch events 'ba; 1' and 'head; 1' both present significant influences. The 'ba; 1'

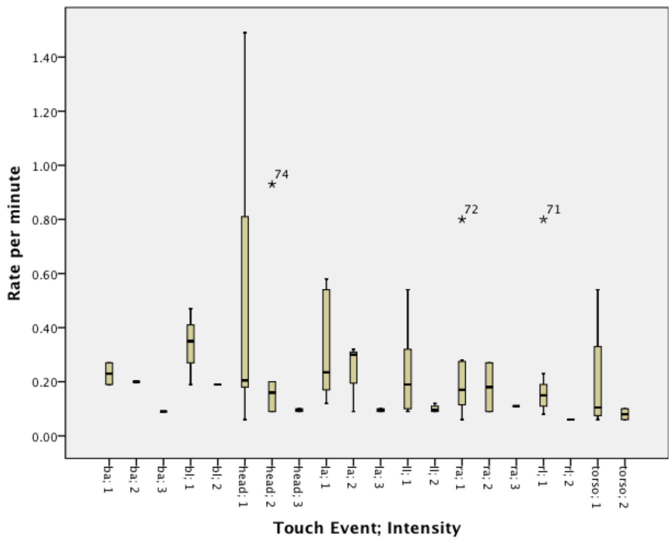


Figure 8. Rate of tactile events based on touch location and intensity.

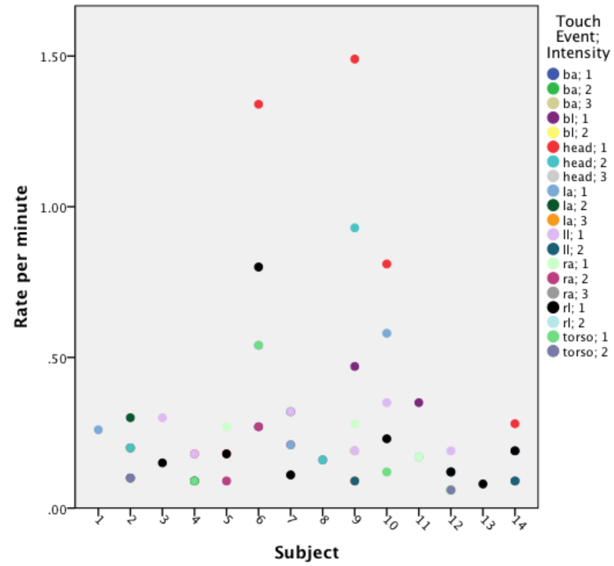


Figure 9. Rate per minute for participant and tactile event; intensity.

indicates slight touch on both robot arms and can be seen on the boxplot to present small variations in rate per minute, while the light head touch identified by 'head; 1' shows stronger variations, thus contributing to the positive slope as reflected by column B.

A. Discussion of results

The visual inspection and statistical analysis of different observed touch types for different participants present a Versatile picture varying across touch types, intensities and different participants. This was predicted prior to the start of this study as touch events are shown to be different in their features, locations and for different participants. This is evident in the statistical results showing significant differences of touch type and participant levels. This poses a potential challenge to taxonomical classification as interactive scenarios with low functioning children with autism often feature free or less-structured interactions. For example, often it is difficult to enforce a certain type of touch for a certain duration and thus studies of the results from such interaction are bound to have large inter/intra participant variations. As shown by statistical results as well as plots presented, the touch rate per minute varies at all factor levels (participant, touch type or touch intensity).

TABLE 4
TOUCH LOCATION AND PARTICIPANT VARIATIONS PRESENTED BY THE GENERAL LINEAR MODEL

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	18.887	1	.000
Participant	104.189	13	.000
TouchEvent	60.008	19	.000

Dependent Variable: Rate per minute
Model: (Intercept), Participant, TouchEvent; intensity

TABLE 5
PARAMETER ESTIMATES FOR THE TOUCH PER MINUTE BASED ON PARTICIPANT AND TOUCH-TYPE; INTENSITY

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.065	.1347	-.199	.329	.231	1	.631
[Subject=1.00]	.135	.1877	-.232	.503	.521	1	.470
[Subject=2.00]	.034	.0864	-.135	.203	.156	1	.693
[Subject=3.00]	.124	.1333	-.137	.386	.868	1	.351
[Subject=4.00]	-.038	.1005	-.235	.159	.144	1	.705
[Subject=5.00]	.026	.0971	-.164	.217	.073	1	.786
[Subject=6.00]	.592	.0924	.411	.773	41.008	1	.000
[Subject=7.00]	.053	.0931	-.130	.235	.319	1	.572
[Subject=8.00]	-.147	.1375	-.417	.122	1.145	1	.285
[Subject=9.00]	.400	.0887	.226	.574	20.340	1	.000
[Subject=10.00]	.241	.0962	.053	.430	6.295	1	.012
[Subject=11.00]	.115	.1200	-.120	.350	.919	1	.338
[Subject=12.00]	-.004	.0890	-.178	.171	.002	1	.968
[Subject=13.00]	-.029	.1786	-.379	.321	.027	1	.869
[Subject=14.00]	0[a]	.	.	.	.	.	.
[TouchEvent=ba; 1]	-.330	.1689	-.661	.000	3.830	1	.050
[TouchEvent=ba; 2]	.101	.1971	-.285	.487	.264	1	.608
[TouchEvent=ba; 3]	-.375	.2062	-.779	.030	3.300	1	.069
[TouchEvent=bl; 1]	.100	.1567	-.207	.407	.410	1	.522
[TouchEvent=bl; 2]	-.275	.2062	-.679	.130	1.773	1	.183
[TouchEvent=head; 1]	.311	.1285	.059	.562	5.844	1	.016
[TouchEvent=head; 2]	.174	.1407	-.101	.450	1.535	1	.215
[TouchEvent=head; 3]	.013	.1632	-.307	.333	.007	1	.935
[TouchEvent=la; 1]	.060	.1408	-.216	.336	.180	1	.671
[TouchEvent=la; 2]	.143	.1500	-.151	.437	.909	1	.340
[TouchEvent=la; 3]	-.187	.1628	-.506	.132	1.314	1	.252
[TouchEvent=ll; 1]	.027	.1285	-.225	.279	.045	1	.831
[TouchEvent=ll; 2]	-.072	.1394	-.345	.201	.269	1	.604
[TouchEvent=ra; 1]	-.010	.1348	-.274	.255	.005	1	.944
[TouchEvent=ra; 2]	-.194	.1706	-.528	.141	1.290	1	.256
[TouchEvent=ra; 3]	-.007	.2085	-.416	.401	.001	1	.972
[TouchEvent=rl; 1]	.045	.1331	-.216	.306	.113	1	.737
[TouchEvent=rl; 2]	-.001	.1971	-.387	.385	.000	1	.995
[TouchEvent=torso; 1]	-.060	.1444	-.343	.223	.173	1	.677
[TouchEvent=torso; 2]	0[a]	.	.	.	.	.	.
(Scale)	.025[b]	.0040	.018	.034			

Dependable Variable: Rate per Minute;
Model: (Intercept), Participant, TouchType
a. Set to zero because this parameter is redundant
b. Maximum Likelihood estimate

 ISBN: 978-1-61208-250-9

This is a true reflection of the captured behaviour and it is in-line with what is confirmed by the literature, i.e., that children with autism are an extremely heterogeneous population and although they share the same core difficulties, each child displays these in an individual way.

This conclusion is also supported by observation analysis of trials where children with autism interacted with the robot in free play context. Figure 10 shows examples of the variety of interaction styles from forceful poking and grabbing (two images on the left) to gentle stroking (image on the right).

In addition, literature suggests that children with autism may demonstrate more interest in parts of objects rather than the object itself as a whole (e.g., wheels on a toy car rather than the car itself) [11]. Sometimes object manipulation is focused on self-stimulatory use of toys rather than exploration of the properties of the object [12].
These behaviours present additional challenges and must be considered when building autonomous robots for cognitive social interaction with children with autism.

V. CONCLUSION AND FUTURE WORK

The paper presented a taxonomical classification of tactile interactions of 14 children with autism with the humanoid robot KASPAR. It described the experiments which were designed to observe the tactile interaction of the children with the robot and record the location and type of these interactions. It than presented the statistical analysis results comparing the observed interaction in term of rate per minutes between different users and different type of tactile interaction. The statistical results presented in the paper showed significant differences across touch type intensities and participants. These results support the literature suggesting that children with autism are an extremely heterogeneous population and although they share the same core difficulties, each child displays these in individual ways. The case study examples discussed in the paper highlight the need and possible benefit for future modular play scenarios and adaptive robots that can be flexible in their operation, i.e., capable of autonomous operation and response to interaction but at the same time allow operation via a remote control where a teacher/therapist could work with the child alongside the robot, triggering additional robot behaviors adapted to the individual needs of the children.

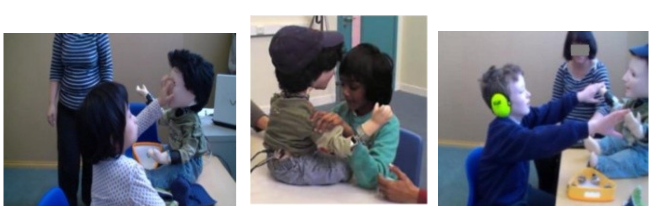


Figure 10. Forceful and gentle interaction of children with KASPAR.

ACKNOWLEDGEMENT

This work has been partially supported by the European Commission under contract number FP7-231500-RoboSkin.

REFERENCES

[1] M. Costonis, *Therapy In Motion.* Urbana: University of Illinois Press, 1978.

[2] P. Bernstein, *Theoretical approaches in dance/movement therapy, I & II.* Dubuque, Iowa: Kendall Hunt, 1986.

[3] O. Bogdashina, *Sensory perceptual issues in autism and Asperger Syndrome: different sensory experiences – different perceptual worlds*. London: Jessica Kingsley Publishers, 2003.

[4] G. Gillingham, *Autism: Handle with Care: Understanding and Managing Behaviour of Children and Adults with Autism*. Arlington: TX. Future Education Inc, 1995.

[5] P. Caldwell, *Getting in Touch: Ways of working with people with severe learning disabilities and extensive support needs*. Brighton: Pavilion Publishing Ltd 1996.

[6] S. Knox, "Development and current use of the revised Knox preschool play scale," in *Play in occupational therapy for children*, Diane Parham and L. Fazio, Eds.: Mosby - Elsevier, 2008, pp. 55-70.

[7] G. Skard and A. Bundy, "Test of Playfulness," in *Play in Coccupational Therapy for Children*, D. Parham and L. Fazio, Eds.: Mosby - Elsevier, 2008, pp. 71-93.

[8] D. Parham and L. Fasio, *Play in Occupational Therapy for Children*. St. Louis: Mosby-Elsevier, 2008.

[9] V. Cumine, J. Leach, and G. Stevenson, *Autism in the Early Years: A Practical Guide* London: David Fulton Publishers, 2000.

[10] K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. Assif Mirza, and M. Blow, "KASPAR- A Minimally Expressive Humanoid Robot for Human-Robot Interaction Research," *Special Issue on "Humanoid Robots", Applied Bionics and Biomechanics*, vol. 6(3): 369-397, 2009.

[11] American-Psychiatric-Association, "Diagnostic and Statistical Manual of Mental Disorders DSM-IV," American Psychiatric Association, Washington DC., 1995.

[12] M. Black, B. Freeman, and J. Montgomery, "Systematic observation of play behaviourin autistic children," *Journal of Autism and Childhood Schizophrenia*, vol. 5, pp. 363-371, 1975.

Resource-Efficient Methods for Feasibility Studies of Scenarios for Long-Term HRI Studies

Nate Derbinsky
Computer Science and Engineering Division
University of Michigan
Ann Arbor, MI, USA
nlderbin@umich.edu

Wan Ching Ho, Ismael Duque, Joe Saunders,
Kerstin Dautenhahn
Adaptive Systems Research Group,
School of Computer Science
University of Hertfordshire
Hatfield, United Kingdom
Email: {w.c.ho,i.duque-garcia2,j.1.saunders,
k.dautenhahn}@herts.ac.uk

***Abstract*—Long term HRI studies can be costly, firstly in terms of researcher time, hardware/software development time, data-collection, data analysis, trial preparation, trial execution, robot time and subsequently, in terms of funding for robotics and other equipment. Methods which reduce such costs by using resource-efficient feasibility studies to analyze study methods and propose outcomes, debug code associated with data collection/analysis, and sanity check human-robot interactions by simulating, predicting and generating feasible scenarios would therefore be welcome. This paper proposes such methods and provides physical implementation details of these methods in practice and data from a preliminary study.**

***Keywords*-feasibility studies; experimental methods**

I. Introduction

Whilst many large research projects propose that humans interacting with robots is achievable given existing robotic technologies and research efforts dedicated to human-robot interaction (HRI), only a relatively small amount of long-term studies have been presented/published in this field. Here long-term studies refer to a series of sustainable experiments that involve one or more human users repeatedly acting/working together with robots over an extended period of time in a complex environment and with a large repertoire of human-robot interaction. Long-term studies are indeed very desirable in investigating various aspects of HRI, such as the key features of a service robot, users' perception and reaction to robots, and scenario and methodological design of HRI, among others. So, in general, what makes the long-term use of service robots in such studies difficult?

One of the major problems for long-term studies of HRI is the resource cost associated with developing such studies. Costs may include researcher time (e.g. developing study materials), hardware and software development time (e.g. building/modifying the robot, coding, data-collection and analysis routines), data-collection time (e.g. preparing for and executing trials), robot time (e.g. time sharing between studies), and funding (e.g. recruiting subjects and purchasing expensive equipment required for studies with complex service robots). To give an example, typically the development of a service robot's functionality (such as the Care-o-Bot 3) in complex interaction scenarios and involving a variety of tasks requires a team of researchers to work on it for several years. If hardware needs to be developed/modified the time frame can be significantly longer until actual real-time studies with users are feasible. If the results of these eventual studies point out flaws in the scenarios that had been implemented, then typically modifications cannot be made easily without again involving significant costs. Such methodological issues have been acknowledged to significantly limit the advancements of the field of HRI ([1], [2], [3]).

Some of these costs may be reduced if resource-efficient feasibility studies can be run to analyze study methods and proposed outcomes, debug code associated with data collection/analysis, and sanity check subject-robot interactions. In our research we are particularly interested in long-term HRI studies, in which we investigate human-robot interactions over weeks and months within a smart-home environment, and it is in these cases that the pay off of feasibility studies could be dramatic.

In this paper, we discuss and evaluate methods that we have been developing to perform resource-efficient feasibility studies in this context. We begin by discussing requirements for these methods. We then introduce methods we have been developing at the University of Hertfordshire Robot House [4], a domestic residential home environment allowing for user-centered HRI scenarios in a familiar and natural context. For each method, we describe the associated goal in terms of reducing the study's resource utilization, analyze the degree to which it meets our methodological requirements, present preliminary evaluation data, and discuss future directions. While these are preliminary methods and findings, we hope to spark a discussion within the HRI community dealing with effective methods for feasibility studies that can improve the efficiency of such methods in our research.

II. Requirements for effective methods

For methods to be effective in feasibility studies for long-term HRI studies, we argue that they must jointly satisfy the requirements of (R1) *resource efficiency* and (R2) *outcome-relative fidelity*.

R1 - Resource Efficiency

The central goal of a feasibility study is to identify relevant issues without expending a great deal of resources. Thus, the method itself must not induce a resource burden. Methods should cheaply, quickly, and broadly produce outcomes that are relevant to the target study.

R2 - Outcome-Relative Fidelity

If a feasibility study is to be useful, its outcomes must be sufficiently accurate and trust-worthy as to support potentially costly design decisions related to further feasibility investigations, as well as the target study. The degree of accuracy necessary will depend upon the types of issues to investigate, but the method should produce outcomes that are informative.

 ISBN: 978-1-61208-250-9

Previous research in HRI had already recognized the need to perform feasibility studies before the completion of the final system. Several methods have been adapted from Human-Computer Interaction to HRI studies, including mock-ups and props [5], video-based [6] and theatre-based [7] methods, and Theatrical Robot [8], or Wizard-of-Oz studies [9] which can be applied once a prototype of the system is available that can be remotely controlled. However, mock-ups for complex robotic systems and their interactions with people are not easy to design, while theatre- and video-based methods lack the situatedness of the user in the interaction context (users are removed from the context by just watching either a video or live-performance). This article presents a method that replicates the situatedness and context of the user embodied in the scenario, despite the lack of a physical embodiment of the robot and its functionalities. Situatedness is an important aspect of embodied cognition and interaction with the physical and social world (e.g. [10]).

III. Methods

In this section we discuss and evaluate three methods we have used for feasibility studies at the UH Robot House.

A. Sensor Logging

Laboratories that design and pursue multiple ongoing HRI studies can quickly find themselves awash in the details of sensors. In addition to installing, testing, and writing software to operationalize sensors, a great deal of time can be spent on trying to unify diverse data formats, unsynchronized timestamps, and decentralized/disorganized log files. To address these concerns, we adopted three main components to our sensor-logging method. First, we centrally log all data to a relational database. The benefits to this approach are numerous: reliability, efficiency, and scalability when dealing with large numbers of sensors, frequent updates, and long-term studies; general data access and query methods, standardized via existing APIs (asynchronous, local or via network) and the SQL query language; synchronization of log values and timing via atomic, consistent, isolated, and durable (ACID) transactions; and flexibility to incorporate arbitrary sensor types/values and associated meta-data.

The second component of our sensor-logging method is sensor registration, which means that we explicitly associate meta-data about sensors (e.g. name, type, location, etc.) with each logged value. This registration is accomplished efficiently via tables and many-to-one relations in the database and can thus be queries later via SQL. This registration process directly facilitates mixed real-virtual sensing; we discuss and exemplify this concept in the Robot Modeling section, but in short, integration of virtual sensors allows rapid prototyping of sensing and analysis algorithms during feasibility testing.

The final component of our sensor-logging methodology is supporting a wide variety of diverse sensors. We use very general database relations for sensor logs and thus provide, in one unified output log, all sensors (real and virtual) that are necessary for Robot House studies. This allows us to capture, in feasibility and target studies, a wide set of phenomena for later analysis and ensures that data read during feasibility studies are representative of those that will be achieved during target studies, with respect to data ranges, precision, etc.

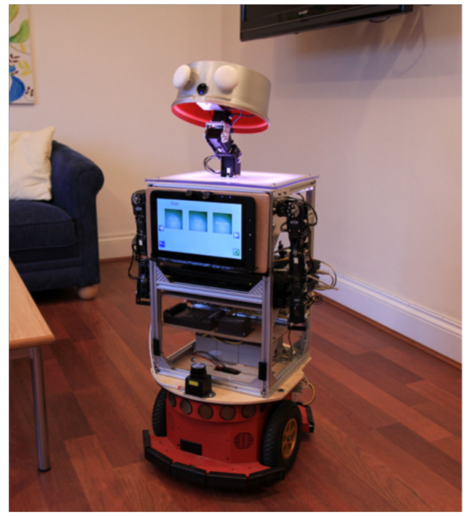


Figure 1. Sunflower robot at the UH Robot House.

1) Evaluation at the UH Robot House: At the UH Robot House, we utilize MySQL [11] as the database management system, which is free, open source, and has an existing ecosystem of software clients and tools. Alongside, two different but complementary commercially available sensor systems, the GEO System [12] and ZigBee Sensor Network [13], are currently installed in the Robot House. The GEO System is a real-time energy monitoring system for electrical devices. It is used to detect the activation and deactivation of electrical appliances by users, e.g. opening the refrigerator, turning on the kettle or the TV, etc. On the other hand, the ZigBee Sensor Network is used to detect those users' activities that cannot be identified by the GEO System such as opening of drawers or doors, occupied chairs or sofa seat places, opening of water taps, etc. In our case, the ZigBee Sensor Network consists of five ZigBee Wireless modules spread across the Robot House. They broadcast all sensors' changes through the Robot House ethernet infrastructure. Each module contains a different number of sensors connected, depending on its location and its use. Both sensory networks together offer a total of 59 sensors to be used in our HRI studies.

Most sensors were used in the Robot House before we introduced this logging method. Therefore, in order to centralize all data coming through both sensor networks, we developed a script connector responsible for updating every 1Hz all the sensors' changes into the aforementioned relational database. We also integrated virtual sensing of tablet touch events, which we will discuss later in the Robot Modeling section. In addition to sensor name and type, we also register location information, which allows us to visualize/analyze sensor logs according to smart-home regions. We have used this method in a single feasibility study with dozens of these sensors, tens of trials, over about an hour of total experimentation time. The default MySQL tuning parameters yielded performance that was sufficient for all of our sensors. During this period we have solidified the database schema and are currently investigating deployment across all studies in the Robot House. We are also investigating how this method can simplify and standardize logging and analysis tools, as well as

 ISBN: 978-1-61208-250-9

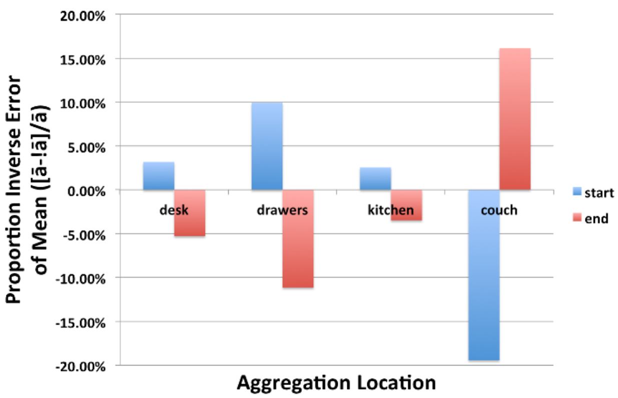


Figure 2. Analysis of navigation model error by location and directionality.

facilitate data sharing, of multiple robot models/smart houses in the ACCOMPANY project [14] across multiple countries/universities.

This method fully satisfies both requirements. Building on existing database technologies and software tools allows us to rapidly (R1) prototype studies as well as analyze experimental data across all sensors at our disposal. Adding a new, or modifying an existing sensor requires minimal time expenditure, allowing rapid prototyping. Also, as the method is ambivalent as to study type (feasibility vs. target) and goals, we can efficiently and seamlessly utilize our system across studies and all sensor logs reflect reported sensor values (R2).

B. Robot Modeling

In laboratories that pursue multiple ongoing HRI studies, robot time can become a scarce resource and time-sharing is required for robot hardware/software development and maintenance, human interaction trials, feasibility studies, etc. To address this issue, we have been investigating limited forms of robot modeling in feasibility studies that is human-robot interaction studies without the robot. While there can be obvious advantages to this approach in terms of resource efficiency (R1), the obvious concern is fidelity of any study outcomes: indeed, any form of simulation/modeling runs the risk of being ”doomed to success” [15]. Furthermore, attempts to develop highly accurate, physics-based, noise-model-driven simulations can exacerbate this problem by being neither resource efficient (R1), nor sufficiently accurate for feasibility studies (R2).

Thus our research approach has been to explore the extreme of resource efficiency, tailored to each specific feasibility study. We ask first, what is the minimal role the robot plays in a particular target study and then we develop methods to minimally model this phenomenon, without a robot. Then, during feasibility testing, we act-out robotic interaction (discussed below) and log interaction events via virtual sensors.

1) Evaluation at the UH Robot House: The broader research goal of our target study was to investigate the role of robot long-term memory in HRI studies. Prior work (e.g. [16] [17]) has shown that long-term episodic memory has the potential to make robotic companions more capable and believable, but there are many open research issues, such as when/what to encode in memory, useful definitions of context for effective retrievals, as well as efficient and scalable algorithms for real-time retrievals over long periods of time [18]. In this context, our target study was to gather long-term episodic traces of HRI studies in the Robot House (on the order of weeks of trial data). Given this data set, one of research goals was to investigate general properties of traces (e.g. data-encoding rates, contextual patterns), as well as task-relevant learning opportunities (e.g. automatic context generation, user preference learning). For this specific goal, and the robot model/smart home, we recognized that we could perform feasibility studies given a small set of human-robot interactions in the context of a set of everyday tasks (e.g. cooking, cleaning, recreating).

We applied our robot-modeling method in two components. First, we investigated the fidelity of modeling two actions (robot navigation and opening/closing of a cargo drawer) of the Sunflower robot (see Figure 1) by observing a small set of trials, fitting a minimal parameterization, and assuming a Gaussian error model. Secondly, we investigated the effectiveness of having the feasibility participant manually log these interaction events via a tablet interface, while acting out the results of the interaction personally (e.g. personally transporting the logging tablet between locations, as opposed to the robot moving). At first glance, these approaches seem almost comically simple but that is the point. These approaches, as discussed below, are incredibly resource efficient, and we gathered data to assess tradeoffs in task-relevant fidelity.

Robot Operation Modeling: For each robot action (navigate-x-to-y, drawer-open/close), we performed three trials with the Sunflower robot in the Robot House and measured operation-completion time (precision = 1 second). For navigation, we used four strategically useful house locations and performed trials at all-pairs, gathering data independently for each direction (12 permutations x 3 trials = 36 data points). We recorded videos for all trials and the data for these models was collected and analyzed in under 1 hour.

Drawer operations executed with no measurable variance in time and symmetrically (i.e. no difference between the time to open and close). Thus, our model for robot drawer operations had zero parameters (2 seconds). We first hypothesized that we could usefully parameterize the navigation operation only by distance traveled. However, the data we gathered showed a poor correlation between the distance the robot traveled and the time required for navigation ($r^2 \approx 0.44$). In practice we could see, unsurprisingly, that more confined spaces caused the robot navigation algorithms to require additional time for correction. Since the locations were not uniform with respect to spatial constraint, we thus recognized the need to parameterize with respect to location.

Based upon this experience, our second hypothesis was that we could usefully parameterize the navigation operation by simply the two endpoints (un-ordered). However, the data showed that the types of navigation corrections necessary were dependent upon directionality: leaving a constrained area took much less time than entering that same area. Figure 2 illustrates our analysis that lead to this outcome, plotting the proportion inverse error of mean time for each location by direction ($[\bar{a}$ - $!\bar{a}]$ / $\bar{a}$, where $\bar{a}$ is average time and $!\bar{a}$ is average time between the same locations in opposite direction). We see in this data that not only is there an asymmetry per directionality in each location, but that asymmetry is not consistent between locations: for the desk, drawers, and kitchen, the error is lower for ending at that location as compared to starting, but this relationship is reversed for the couch (which has a table nearby). Consequently, our navigation model required start location, end location, and directionality.

Given these three parameters, three measurement trials yielded a relatively high variance (up to 35% of total operation time). For our feasibility study, this degree of accuracy was sufficient to gather relevant data. However, other studies may require tighter bounds, which may entail additional measurement trials and/or more sophisticated error models (leading to additional fidelity vs. resource-efficiency analysis).

Robot Interaction Modeling: To gather the data we required, it was necessary not only to know how long robot actions would take, but log these events in context of other Robot House sensor values while subjects were engaging in various scripted activities. This required (1) integrating robot event logs with other sensor inputs and (2) triggering these events at appropriate times without using an actual robot.

First, we registered a virtual sensor in our sensor-logging platform (see above). To trigger this sensor, we built an HTML5, touch-enabled web interface (see Figure 3) using the iUI framework [19]. The front-end interface (Figure 3: top) was arbitrarily extendable to support many robot actions and, upon submission, a PHP script would generate timing data based upon the robot operation model and send the data to our logging platform via HTTP. After logging the event, the interface displays confirmation and a countdown of operation time (Figure 3: bottom). Subtly, the physicality of the device used and this countdown provides feedback to the user of when a robot would be performing the action, and might thus be not available for interaction. The benefits of the technologies we used are rapid development; intuitive, touch interface for subjects; as well as platform- (Windows, *nix, Mac) and device- (desktop/laptop, tablet, phone) independence.

The source of the event trigger is a complex decision, and for preliminary evaluation we opted to evaluate the simplest choice: the subject would act-out the robot him/herself. Thus, if the subject wished for the robot to take cargo (e.g. a plate) to a destination (e.g. the kitchen), s/he would perform the following sequence of actions: log a ”drawer open” action (wait for countdown), log a ”drawer close” action (wait for countdown), log a ”navigate” action, bring the interface device and cargo to the destination, leaving both (and not having further interaction with the interface till countdown completed). As expected, this approach places numerous cognitive and physical burdens on the subject, such as remembering to initiate robot actions [in proper order], maintaining task priorities, and actually performing actions. This approach also affects fidelity (R2) in that the subject cannot simultaneously perform a robot action and a study action. However, the resource cost of the feasibility study is now limited to a single individual and no robot (R1). In our feasibility-study trials, we found that even a well-trained subject was likely to make numerous event-logging errors (e.g. forgetting to log an event), and thus in the future we plan to evaluate a slightly more complex act-out option: utilizing a second person as an independent actor (sacrificing R1 for R2).

C. Interaction Script Generation

In designing long-term HRI studies, a great deal of time can go into developing user-behavior scripts. Some of the complexities

Figure 3. Robot-interaction and virtual-sensor interface (sequence of screenshots for an action on three different devices/programs).

involved are reflecting the appropriate level of abstraction in subject tasks (reflecting necessary task structure, while not inhibiting how subjects individualize their execution); controlling for sources of stochasticity in task organization and execution; as well as scaling to large numbers of participants over long periods of time. We believe that feedback from rapid feasibility studies can improve the overall time to develop user-behavior scripts, and have focused on a flexible, intuitive framework to describe user activities and sources of randomness, as well as quickly and reproducibly produce scripts according to this framework.

Before we describe our method, we note that the problem of producing user-behavior scripts has many elements in common with controlling non-playing-character (NPC) behavior in video games (e.g. [20] [21]), assuming individual users' behavior is driven by simplified motivational states. For simple NPCs, finite state machines (FSM) can fully describe behavior. However, for more complex characters (or units of characters) in interesting environments, it becomes too complex to develop and maintain a step-by-step characterization of every suitable action in every state (especially for non-programmers) [21]. However, on the opposite extreme, it is often too computationally expensive to simply utilize off-the-shelf planners during real-time play [20]. Thus, one reasonable approach to this problem is to have an intuitive, often graphical framework by which to hierarchically describe goals and organizations of contexts (e.g. via Hierarchical Task Networks, or HTNs); efficiently instantiate a script based upon this organization; and then rely upon computationally efficient algorithms to implement script primitives. At a high level, this describes our method for user-behavior script generation.

Our hierarchical description language draws inspiration from HTNs: an acyclic graph where internal nodes are used for or-

ISBN: 978-1-61208-250-9

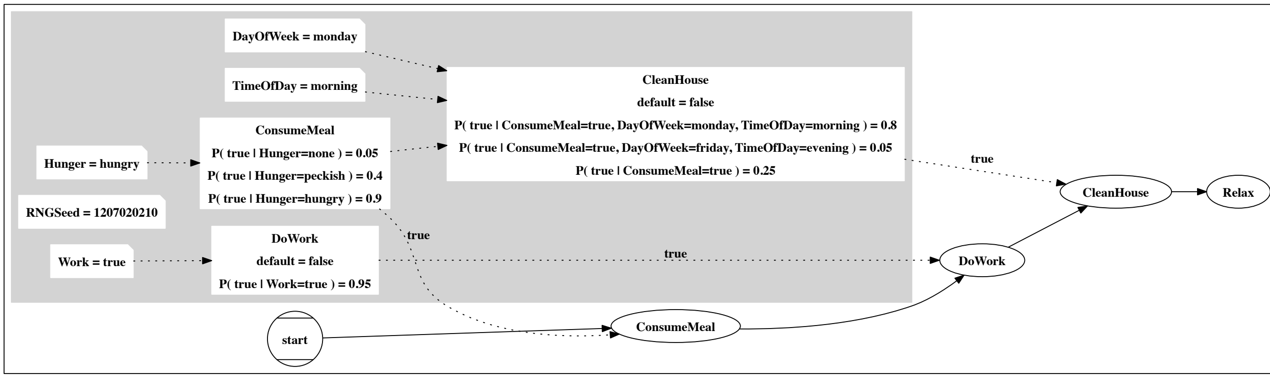


Figure 4. Example top-level script-structure description.

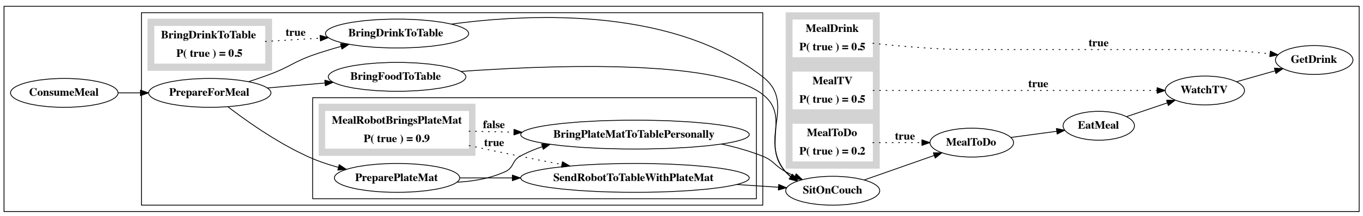


Figure 5. Expansion of Figure 4, showing substructure of the ConsumeMeal node.

ganization and leaf nodes indicate behavior primitives. However, drawing on Bayesian models for inspiration, we also support describing probabilistic dependencies. Thus, each internal node supports an arbitrary set of variables, and each edge supports a conditional-probability table (CPT), describing uniform distributions over actions. For example, consider Figure 4, an example used at the UH Robot House that describes a high-level sequence of user behaviors consisting of consuming a meal, doing some work, cleaning the house, and relaxing. The greyed area shows variables, variable dependencies, variable values (for a particular script instantiation), and conditional-probability distributions of performing actions based upon variable values. For example, there is a 90% chance that the subject will be instructed to consume a meal (and, if so, an 80% chance that the subject will be instructed to clean the house). Each of the nodes supports arbitrarily deep recursive structure, depending upon the needs of the study (e.g. see a possible expansion of the ConsumeMeal node in Figure 5). Given this high-level description, it is possible to efficiently generate an arbitrarily large set of output scripts that abide by the structure (by controlling input variables and random-seed generation).

1) Evaluation at the UH Robot House: We implemented the high-level description language as a Java library. To generate scripts we implemented a hierarchical-decision agent using the Soar cognitive architecture [22] [23]: the Java library compiles the graph, inputs it to the agent (along with requisite variable values and, optionally a random seed to control stochastic decisions), and receives a sequence of leaf nodes as a script-agent output. We also implemented visualization routines using Graphviz [24] for debugging and communication purposes (e.g. see Figures 4 and 5).

In our feasibility studies, we wrapped the above library/agent within a client program that requested as input all information required to uniquely identify a subject trial (e.g. date, trial #) such that the output script could be later regenerated for examination or as input to learning/activity-recognition algorithms. The script generation program takes a fraction of a second to run for a graph with dozens of nodes and variables. For example, here is a sample input/output based upon the graphs described in Figures 4 and 5:

```
Input> year=2012 month=7 day=5
       time=morning hunger=peckish
       work=true trial=3

Output> start ->
        SendRobotToTableWithPlateMat ->
        BringFoodToTable ->
        BringDrinkToTable -> SitOnCouch ->
        EatMeal -> GetDrink ->
        TurnOnMusic -> WorkToDo ->
        WorkForTenMinutes ->
        GetPeriodicalFromDrawers ->
        LieOnCouch -> RelaxToDo -> Read
```

This example illustrates how an intuitive, hierarchical description of a complex, stochastic (but structured) study can efficiently result in a script that has various degrees of specificity. For example, where necessary there are very concrete actions (e.g. sending the robot to the table with the plate mat), but also those with a great deal of flexibility left to the subject (e.g. bring food to the table). In one feasibility study, we used this software to generate a suite of plans (3 parameter settings, 2 trials each, for 1 subject), all of which in about 1 minute of time (R1).

To facilitate fast and accurate (R2) development, in the future we plan to develop a GUI to support drawing of graph nodes and visually entering CPT values for non-programmers.

IV. CONCLUSIONS

In this work, we have described and evaluated methods to perform HRI feasibility studies that can dramatically reduce resource utilization when compared to full target studies, but maintaining the high level of HRI fidelity. We have incorporated all of these methods in our work at the UH Robot House. Our sensor logging approach has clearly demonstrated effectiveness and we are beginning deployment across all of our studies, while the other methods require additional investigation.

The goal of this paper was to begin a conversation in the HRI community around the value of feasibility studies for complex HRI scenarios and the need to investigate and disseminate software and lessons-learned from practical techniques therein. We hope that future work highlights innovations in this space, which we contend will improve the speed and quality with which studies can be executed in this field.

ACKNOWLEDGMENT

The first author thanks the National Science Foundation for funding an extended research visit to the UH Robot House, without which this research would have not been possible. The work described in this paper was partially supported by the EU Integrated Project ACCOMPANY (Acceptable robotiCs COMPanions for AgeiNg Years) [14]. Funded by the European Commission under contract number 287624.

REFERENCES

[1] Michael L. Walters, Sarah Woods, Kheng Lee Koay, and Kerstin Dautenhahn. Practical and methodological challenges in designing and conducting interaction studies with human subjects. In *University of Hertfordshire, UK*, pages 110–120, 2005.

[2] Kertin Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):679–704, 2007.

[3] Kerstin Dautenhahn. Methodology and Themes of Human-Robot Interaction: A Growing Research Field . *International Journal of Advanced Robotic Systems*, 4(1):103–108, 2007.

[4] University of Hertfordshire (UH) Robot House. http://www.hertsrobots.com/. [Online; accessed 12.12.2012].

[5] Christoph Bartneck and Jun Hu. Rapid prototyping for interactive robots. In *The 8th Conference on Intelligent Autonomous Systems (IAS-8)*, pages 136–145. IOS press, 2004.

[6] Michael L. Walters, Manja Lohse, Marc Hanheide, Britta Wrede, Kheng Lee Koay, Dag Sverre Syrdal, Anders Green, Helge Hüttenrauch, Kerstin Dautenhahn, Gerhard Sagerer, and Kerstin Severinson-Eklundh. Evaluating the robot personality and verbal behaviour of domestic robots using video-based studies. *Advanced Robotics*, 25:2233 – 2254, December 2011.

[7] Amiy R. Chatley, Kerstin Dautenhahn, Mick L. Walters, Dag S. Syrdal, and Bruce Christianson. Theatre as a discussion tool in human-robot interaction experiments - a pilot study. In *Proceedings of the 2010 Third International Conference on Advances in Computer-Human Interactions*, ACHI '10, pages 73–78, Washington, DC, USA, 2010. IEEE Computer Society.

[8] Ben Robins, Kerstin Dautenhahn, and Janek Dubowski. Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction Studies*, 7(3):509–542, 2006.

[9] Anders Green, Helge Hüttenrauch, and Kerstin S. Eklundh. Applying the wizard-of-oz framework to cooperative service discovery and configuration. In *Proceeding of the 13th IEEE International Workshop on Robot and Human Interactive Communication, ROMAN04*, pages 575–580, Kurashiki, Okayama, Japan, September 2004.

[10] Andy Clark. *Being There: Putting Brain, Body, and World Together Again*. The MIT Press, January 1998.

[11] MySQL. http://www.mysql.com/. [Online; accessed 12.12.2012].

[12] GEO: Green Energy Options. http://www.greenenergyoptions.co.uk/. [Online; accessed 12.12.2012].

[13] Shahin Farahani. *ZigBee Wireless Networks and Transceivers*. Newnes, Newton, MA, USA, 2008.

[14] ACCOMPANY: Acceptable robotiCs COMPanions for AgeiNg Years. Project No. 287624. http://www.accompanyproject.eu/, October 2011. [Online; accessed 12.12.2012].

[15] Rodney A. Brooks and Maja J. Mataric. Real robots, real learning problems. In Jonathan H. Connell and Sridhar Mahadevan, editors, *Robot Learning*, volume 233 of *The Kluwer International Series in Engineering and Computer Science*, pages 193–213. Springer US, 1993.

[16] Paulo F. Gomes, Carlos Martinho, and Ana Paiva. I've been here before! location and appraisal in memory. In *Proc. of The 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*. IFAAMAS (http://www.aamas-conference.org/), 2011.

[17] Mei Yii Lim, Ruth Aylett, Patricia A. Vargas, Wan Ching Ho, and João Dias. Human-like memory retrieval mechanisms for social companions. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 3*, AAMAS '11, pages 1117–1118, Richland, SC, 2011. International Foundation for Autonomous Agents and Multiagent Systems.

[18] John E. Laird and Nate Derbinsky. A year of episodic memory. In *Proceedings of the Workshop on Grand Challenges for Reasoning from Experiences*, pages 7–10, Pasadena, CA, USA, July 2009.

[19] iUI. http://www.iui-js.org/. [Online; accessed 12.12.2012].

[20] John-Paul Kelly, Adi Botea, and Sven Koenig. Offline planning with hierarchical task networks in video games. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Interactive Digital Entertainment AIIDE-08*, pages 60–65, 2008.

[21] Antonio A. Sánchez-Ruiz, David Llansó, Marco A. Gómez-Martín, and Pedro A. González-Calero. Authoring behaviours for game characters reusing automatically generated abstract cases. In Sarah Jane Delany, editor, *ICCBR Workshop*, pages 129–137, July 2009.

[22] Soar v9.3.2. http://sitemaker.umich.edu/soar/. [Online; accessed 12.12.2012].

[23] John E. Laird. *The Soar Cognitive Architecture*. MIT Press, 2012.

[24] Graphviz. http://graphviz.org/. [Online; accessed 12.12.2012].

Person Identification using Skeleton Information from Kinect

Aniruddha Sinha
Innovation Labs,
Tata Consultancy Services
Kolkata, India
aniruddha.s@tcs.com

Kingshuk Chakravarty
Innovation Labs,
Tata Consultancy Services
Kolkata, India,
kingshuk.chakravarty@tcs.com

Brojeshwar Bhowmick
Innovation Labs,
Tata Consultancy Services
Kolkata, India
b.bhowmick@tcs.com

***Abstract*—In recent past the need for ubiquitous people identification has increased with the proliferation of human-robot interaction systems. In this paper we propose a methodology of recognizing persons from skeleton data using Kinect. First a half gait cycle is detected automatically and then features are calculated on every gait cycle. As part of new features, proposed in this paper, two are related to area of upper and lower body parts and twelve related to the distances between the upper body centroid and the centriods derived from different joints of upper limbs and lower limbs. Feature selection and classification is performed with connectionist system using Adaptive Neural Network (ANN). The recognition accuracy of the individual people using the proposed method is compared with the earlier methods proposed by Arian et. al and Pries et. al. Experimental results indicate that the proposed approach of simultaneous feature selection and classification is having better recognition accuracy compared to the earlier reported ones.**

Keywords-Person identification; gait recognition; adaptive artificial neural network(ANN); Kinect; connectionist system

I. INTRODUCTION

Automatic person identification is one of the important factors in Human-robot interaction based applications [1]. Successful identification of individual gait pattern can certainly help a machine to take and control different actions on the basis of previous interaction with that same individual. The interaction by the robot can be personalized upon identification of an individual [2][3]. Biometric identification, using different behavioral patterns like gait, keyboard typing, lip movement etc. or physiological signatures like voice/speech, face, iris, fingerprint etc. [4], is one of the common means of automatic person recognition. Over the last few decades, extensive research work has already been done on different biometric modalities like face, iris scan and fingerprint recognition [5]. However, other biometric characteristics like gait, skeleton data, EEG are relatively less considered [8]. Moreover, it is very difficult to get iris, fingerprint, face or audio (voice/speech) related biometric information (at the recognizable resolution) from a large distance and without user's direct co-operation. Human gait recognition has great advantages in recognition from low resolution images where other biometric techniques are not suitable because of insufficient pixel information [6]. Psychophysical studies [7] indicate that human being is capable of recognizing an individual reliably, from his/her style of walking or movement. Another advantage of gait pattern identification is that it is very much hard to hide or imitate.

For gait information processing, we need video sequence of walking person, where at least one complete gait cycle is present. A gait cycle is starting with one foot forward and ending with same foot forward. Each and every gait has two main components [9] namely, a structural component or physical build of a person (e.g. height, length of limbs etc.) used to derive static features and motion dynamics during gait cycle used to derive dynamic features.

In this paper, we use Microsoft Kinect sensor to derive the skeleton data for recognizing of individuals. The main contribution of this paper is given below:

- Automatic gait cycle detection algorithm from skeleton data of Kinect as opposed to manual gait cycle detection [22].
- Proposal of new static and dynamic features related to area and distance.
- Methodology to use ANN based connectionist framework for feature selection and supervised learning algorithm [30] to determine the important features.

The proposed area features include the area spanned by the lower and upper parts of the body as physical/structural features. The distance features include relative distances between body-centroid and the centroids formed by the joints of upper and lower limbs. Comparison of recognition accuracy is done between the proposed approach and the one mentioned in [21] and [22].

This paper is structured as follows: Related work is presented in Section II. Proposed method for Kinect data recording setup and feature extraction are given in Section III. The supervised learning and ANN based feature selection is given in Section IV. The results are given in Section V followed by conclusion and acknowledgement.

II. Related work

Existing gait recognition methods can be categorized under two categories namely model based and model free approaches [10]. In the model based approach, gait signature is derived from modeling and tracking different parts of the body (like legs, limbs, arms etc) over time and then it is used for person recognition and identification. Primary model based approaches consider only static parameter for recognition [11]. BenAbdelkader et al. [12] used structural

 ISBN: 978-1-61208-250-9

parameter like stride and cadence, of walk-cycle for identification. Yam et. al [13] built a structure and motion based model of legs to indiscriminate walking and running gait pattern using biomechanics of human and pendulum motion. Though this approach is view-invariant and scale independent, but it is computationally expensive and very much sensitive to quality of gait sequences [10]. Model free approaches mainly characterize gait pattern by observing how shape of the silhouette of an individual changes over time or by considering entire motion dynamics of the person. Sarkar et. al [14] propose a baseline algorithm using series of gait silhouettes as feature. Gait Energy Image (GEI) [15] and Motion Energy Image (MEI) [16] are also used for individual identification. Though GEI is robust but it lacks dynamic gait information. Liu and Zheng [17] propose an algorithm which combines both spatial and temporal information. To solve the problem of gait incompleteness, Chen et. al [18] suggests FDEI algorithm. Xue et al. [19] use wavelet decomposition of GEI for infrared gait recognition. In many of these approaches, width of the outer silhouette contour has been considered as a good feature candidate. Model free approaches are relative easy but not robust with view point and scale. For direct gait classification, many researchers use KNN, SVM and TSVM. Dynamic Time Wrapping (DTW) is a very popular technique for gait pattern matching. Sundaresan et al. [23] successfully develop a HMM based framework for gait recognition. Recently [21] has employed 3D skeleton based approach for recognizing individual person from their walking gait pattern. They have used Microsoft Kinect to capture skeleton data and used K-means as an unsupervised learning algorithm for clustering. Pries et. al. [22] also does a comparable job of skeleton based person recognition using Kinect. They have used 11 static and 2 dynamic features for classification and compared 1R, C 4.5 decision tree and Naive Bayes classification results. To best of our knowledge there have been no work done on the feature selection of Kinect data for person identification. PCA [20] and LDA [24] are most commonly used feature selection tool in this field. More recently, Mu and Tao [25] use DLA to reduce biologically interdependent features. In this paper we investigate the importance of various features using the ANN based feature selection method.

III. PROPOSED METHOD

In this paper we propose a skeleton based approach (provided by Microsoft Kinect Sensor) for gait recognition and person identification. Kinect provides real time human skeleton information of 20 skeleton joints. We have used Microsoft Kinect SDK for recording the data for human skeleton at 30 frames per second (fps).

Our system is implemented mainly in three steps - recording of skeleton data for side walking pattern using Kinect (section III.A), feature generation over half gait cycle (section III.B) and supervised learning and feature selection to identify an individual (section III.C).

A. Skeleton Data Recording

We recorded the 20 joints of skeleton data for a person, keeping position of Kinect fixed throughout our experiment. In our experiment, side way walking pattern of an individual on an arbitrary path is considered. In each recording session, we have taken multiple side-walks (from left to right and right to left), where the distance of the persons from the Kinect ranged between 6 feet to 10 feet (Figure 1). This enables us to test the identification accuracy in a scenario where the distance of the subject changes over time.

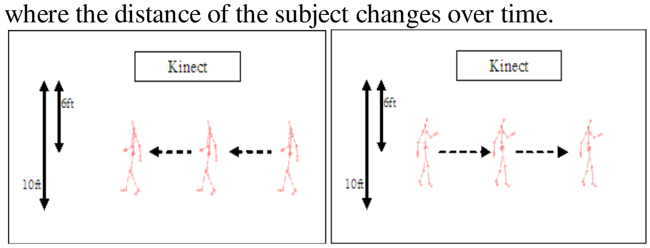


Figure 1. Recording setup using Kinect -- (a) showing right to left walk and (b) showing left to right walk

B. Feature Generation

Salient feature generation from skeleton data is an important step to discriminate individual characteristics. In case of walking, half gait cycle provides all the meaningful information required for identification. Thus initially, the half gait cycles need to be identified from the skeleton data and then generate the features per half-gait cycle.

Computation of Half-Gait Cycle

The side walk movement pattern of a person is shown in Figure 1. We have considered x-axis as horizontal axis, parallel to ground. Figure 2 shows the changes in left and right ankle position over time (frame number) indicating an individual is moving multiple times from left to right and right to left in front of the Kinect.

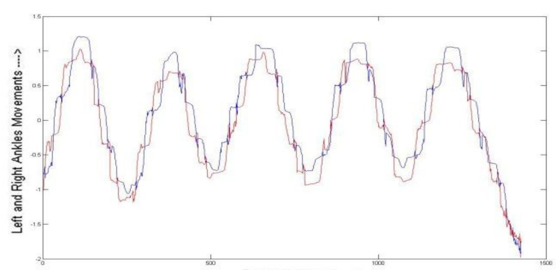


Figure 2. Left (marked in Red color) and Right (marked in Blue color) Ankle movement profile

Horizontal distance i.e., difference between x-co-ordinates of left and right leg ankles is computed for each recorded frame using (1).

$$distance_k = \left| x_{left_{ankle}} - x_{right_{ankle}} \right| \quad (1)$$

For all k = [1, N], where N = total number of frames for individual sidewalk (N > 1).

This distance vector is smoothed using moving average algorithm with small window size. We have detected the transition from negative slope to positive slope to identify all the local minima. The half gait cycle is considered as the frames between two consecutive local minima. Figure 3 shows the square of $distance_k$ for several frames, in dotted green color and the half gait cycle boundaries are shown with vertical partitions in red color.

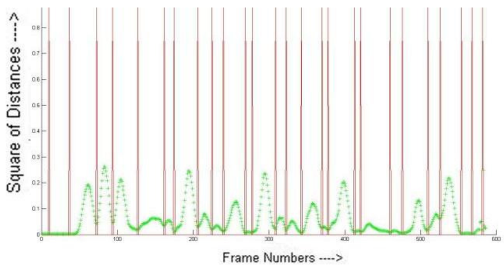


Figure 3. Detection of Half Gait Cycle

Feature Vector

The features that can be generated from the skeleton data of 20 joints can broadly be classified as static and dynamic features. The static features are based on the distances between the adjacent physical joints as shown in the following equation

$$d_{ij} = \sqrt{||x_i - x_j||^2 + ||y_i - y_j||^2 + ||z_i - z_j||^2}$$

where the joints i and j are directly connected and adjacent to each other. The dynamic features include the variation due to the joint movements and their orientations with respect to the other joints which are not adjacent to each other.

In [21] Adrian et. al have proposed eighteen features related to the angle changes in different joints below hip. These are mainly dynamic features where they have taken the mean, max and standard deviation of the three angles for each leg. We have considered these features for comparing the accuracy of individual identification. In [22] Preis et. al have proposed 12 static features (distances between joints) and 2 dynamic features (stride length and speed) for people identification. These are also considered for comparison.

In this paper we have proposed a two area related features and a set of hybrid features which are a mixture of static and dynamic information. Out of these, the first area feature f_{au} is static information of an individual and changes insignificantly while walking or movement. The second area feature f_{al} is a dynamic feature and changes significantly while walking. The proposed twelve hybrid features are distance features f_D which are related to static and dynamic information of an individual.

Area Features: Area occupied by upper and lower part of the body during side walk is the key distinguishing factor as the spread of hand and leg for each person is different from others. So we have considered area of upper and lower body during side walk as one of our features.

A reasonable way to compute area of upper or lower part of a body is to select N joints ($3 \leq N \leq 20$), so that N joints form a closed polygon. If co-ordinate of i^{th}($\forall\, i \in N$) joint is (x_i,y_i), then the area enclosed by N joints, as projected on Z-plane, is

$$A = \frac{1}{2}\sum_{i=0}^{N}(x_i * y_{i+1} - x_{i+1} * y_i) \quad (2)$$

Using (2) we have calculated area of upper body (f_{au}) and lower body (f_{al}). The joint considered for upper body area are shoulder centre, shoulder left, hip left, hipcentre, hip right and shoulder right. The joints considered for lower body area are hip centre, hip right, knee right, ankle right, ankle left, knee left and hip left. The final area feature vector is the mean area values in a half-gait cycle, given as $f_A = \{f_{au}, f_{al}\}$ is in R^2 space.

Distance Features: The change in distances of skeleton joints with respect to the centroid of the upper body is unique for a person and is easily recognizable by a human brain. Thus, the Euclidean distances between centroid of different parts of the body with respect to the centroid of the upper body are indeed a set of good candidate as features. Here, we have considered only four distances between upper body centroid and to the centroids of both hands and legs, separately.

For a closed polygon of N vertices, the centroid $\overrightarrow{X_c}$can be computed as

$$\overrightarrow{X_c} = \frac{1}{N}\sum_{i=0}^{N} N(x_i, y_i, z_i) \quad (3)$$

Using (3), we have calculated the centroid for the following:

- upper body (enclosed by shoulder centre, shoulder left, hip left, hip right and shoulder right)
- right hand (enclosed by shoulder right, elbow right and wrist right)
- left hand (enclosed by shoulder left, elbow left and wrist left)
- right leg (enclosed by hip right, knee right and ankle right)
- left leg (enclosed by hip left, knee left and ankle left)

The Euclidean distances f_{ci} between upper body centroid (x_c, y_c, z_c) and other centroids (x_i, y_i, z_i) can be written as given in (4) where, i=1 for left hand, i=2 for right hand, i=3 for left leg and i=4 for right leg.

$$f_{ci} = \sqrt{||x_c - x_i||^2 + ||y_c - y_i||^2 + ||z_c - z_i||^2} \quad (4)$$

$$\forall\, i \in \{1,2,3,4\}$$

Thus the final hybrid distance feature vector is in R^{12}spacegiven by (5)

$$f_D = \{f_{ci}^{mean}, f_{ci}^{stddev}, f_{ci}^{max}\} \quad (5)$$

where, $f_{ci}^{mean} = mean(f_{cik})$,$f_{ci}^{stddev} = stddev(f_{cik})$, $f_{ci}^{max} = max(f_{cik})$, $\forall k \in halfgait$

Figure 4 shows Euclidean distance between upper body centroid and right hand centroid.

We have also considered all the features ($f_{[21]}$) and ($f_{[22]}$) reported in the papers [21] and [22] respectively however, instead of taking angle between foot and ground, angle between ankle and ground is used because left and right foot co-ordinates are relatively more inclined to noise. Thus finally the combined feature vector is in R^{46} space as given by (6).

 ISBN: 978-1-61208-250-9

$$f = \{f_{[21]},\ f_A,\ f_D,\ f_{[22]}\} \quad (6)$$

The feature data are the normalized to zero mean and unit co-variance across all the features before using for training and recognition.

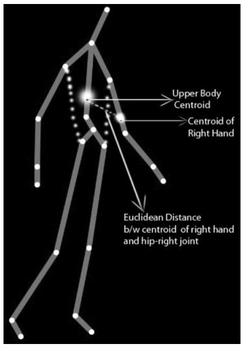


Figure 4. Euclidean distance between upper bodycentroid and right hand centroid

IV. SUPERVISED LEARNING

We have used Adaptive Neural Network (ANN) and Naive Bayes classifier for supervised learning. Pries et. al. in [22] have found that Naive Bayes performs better compared to 1R and C4.5 decision tree. Hence we have excluded them in our experiment.

A. Naive Bayes

The Naive Bayes classifier is based on Bayes' law. It is a probabilistic classifier which assumes the statistical independence of the features. This assumption is violated in the case of features extracted from the human gait cycle as clearly indicated by [22], however, it performs the best as reported with the dataset tested in [22]. Thus we also have considered the Naive Bayes classifier for comparing the proposed features along with one more non-linear classifier.

B. Multi layer Perceptron Models

Multi-layer Perceptron Model (MLP) is a well know ANN architecture [26]. This consists of an input layer, an output layer and one or more hidden layers. The excitation function of the neurons (nodes) used for the experiment is a uni-polar sigmoid function. The strengths of the connections between the nodes of different layers are termed as weights. These weights are initialized randomly and the final weights are obtained after running the back-propagation (BP) algorithm [27] iteratively over the training feature vectors. A sample structure of the ANN is shown in Figure 5 where a single hidden layer is shown but in general multiple such hidden layers can be present. The weight matrix $\overrightarrow{W^1}$ corresponds to the weights between the hidden layer and the input layer. Similarly, $\overrightarrow{W^2}$ corresponds to the weight matrix between the hidden layer output layer and the output layer. The output of the neural network is denoted by the hypothesis function $h_w(f)$.

In our experiments we have considered only a single hidden layer. It is observed that increasing more number of nodes or increasing number of hidden layer is not improving the results and it was taking more time to execute. By experimentation and using the method proposed by Hagiwara [29], we have chosen a single hidden layer with number of nodes as 25.

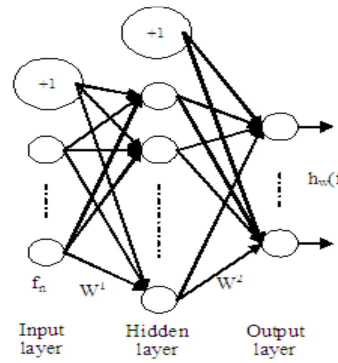


Figure 5. A sample structure of the ANN

The cost function of the neural network is given by

$$J(w) = \frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}[-y_k^{(i)}\log((h_w(f^{(i)})_k) - (1-y_k^{(i)})\log(1-h_w(f^{(i)})_k)] + \frac{\lambda}{2m}[\sum_{j=1}^{H}\sum_{k=1}^{N}(W_{jk}^1)^2 + \sum_{j=1}^{K}\sum_{k=1}^{H}(W_{jk}^2)^2] \quad (7)$$

where various terms are described below:

- m is the number of training samples
- The input training features are $f_n^{(i)}$, where n = 1, 2 ... N and i = 1, 2, ... m. Here N is the number of features and m is the number of training samples. Thus is $f_n^{(i)} \in R^N$ an N dimensional feature vector.
- K is the number of classes which is the number of nodes in the output layer. The class corresponding to the feature vector is a vector $y_k^{(i)}$ where k = 1, 2, ... K. For the class vector corresponding to p^{th} class is (0, 0, .. 1, .. 0) where the p^{th} entry is 1 and remaining are 0.
- H is the number of hidden layer.
- λ is the regularization parameter
- $\overrightarrow{W^1}$ and $\overrightarrow{W^2}$ are the weight matrices.

The steps for training the neural network is given in [30].

We have used K-fold cross validation to choose regularization parameter λ. Figure 6 shows the validation error with respect to λ.

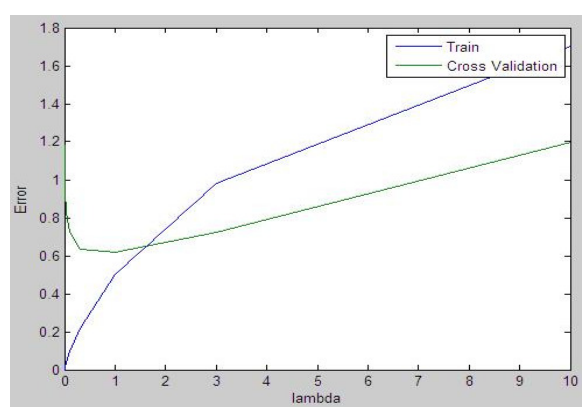


Figure 6. Validation curve of the ANN

The regularization parameter λ is selected as 0.6 which gives the minimum cross-validation error. Learning curve is used to verify that the training samples are enough to train the designed ANN structure (Figure 5).

C. Feature Selection using ANN

In this paper, we provide a methodology to use the connectionist system for feature selection as proposed by Pal et. al [30] in order to get more insight to the previous and new features. In [30], an "on-line" methodology is provided for simultaneous feature selection and classification using MLP. We have used the same approach; however the learning is done in "Batch" mode rather than "on-line" mode. For each iteration, in the "Batch" mode, all the training samples are fed to the neural network and the error in each layer are accumulated to compute the gradient values of the cost function with respect to the weights. Whereas in "on-line" mode, for each iteration, a single training sample is used to compute the error in each layer to compute the gradient values. After every iteration, the weights are updated using conjugate gradient descent algorithm to minimize the cost function. The ANN structure is shown in Figure 7 where a "Selection Layer" is added before the standard input layer.

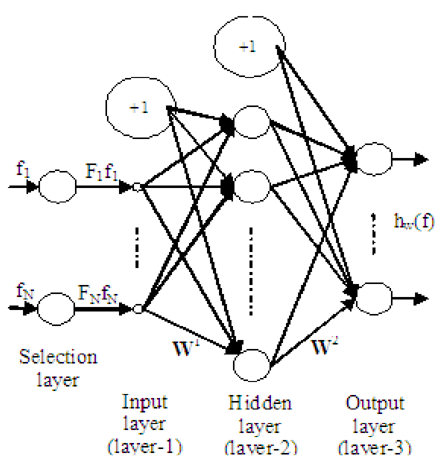


Figure 7. ANN Structure with Feature Selection

According to [30], the attenuation function in the selection layer is F and the argument for the attenuation function is M_n for the n^{th} feature. This attenuation function F is applied to the input features before passing to the normal neural network. During the learning process the arguments M_n gets updated such that the favourable features having higher discriminative properties get higher values in $M_n$while others get lower values. Thus the favorable features, termed as good features, get passed with higher strength while the others get lower strength. This process is called the selection of features which happens simultaneously with the training of the ANN.

In our experiment the attenuation function is chosen as a uni-polar sigmoid function given by (8).

$$F(M) = F_n = \frac{1}{1+e^{-M_n}} \tag{8}$$

Thus the modified input to the input layer is $F_n f_n$ for n = 1,2 ... N, where N is the number of input features.Thus the modified cost function for the batch mode is given by (9), where, $\overrightarrow{f_a^i} = \overrightarrow{f^i} * \overrightarrow{F(M)}$ is the i^{th}attenuated feature vector for the i^{th} training sample.

$$J(W,M) = \frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}[-y_k^{(i)}\log((h_w(f_a{}^{(i)})_k) - (1-y_k^{(i)})\log(1-h_w(f_a{}^{(i)})_k)] + \frac{\lambda}{2m}[\sum_{j=1}^{H}\sum_{k=1}^{N}(W_{jk}^1)^2 + \sum_{j=1}^{K}\sum_{k=1}^{H}(W_{jk}^2)^2] \tag{9}$$

The gradients of the cost function (as in (9)) with respect to the weights ($\frac{\partial J}{\partial W^1}and\frac{\partial J}{\partial W^2}$) and with respect to the arguments for the attenuation function ($\frac{\partial J}{\partial M}$)are computed using BP algorithm. In every iteration, all the training samples are used to accumulate the gradient values. These gradients are then used to calculate the updated weights and arguments for the attenuation function.

The derivation for updating$\overrightarrow{W^1}$, $\overrightarrow{W^2}$and $\overrightarrow{M}$ are given in [30] which are based on on-line mode. For the batch mode, the gradient vectors in each layer are accumulated to compute the gradient of the cost function. The steps for updating the $\overrightarrow{W^1}$, $\overrightarrow{W^2}$and $\overrightarrow{M}$ are given below:

1. Compute the feed-forward for the feature vector of the i^{th} training sample $\overrightarrow{f^{(i)}}$.
2. Calculate the error δ_k^3 in for each output node k in layer-3 by setting $\delta_k^3 = h(f^{(i)})_k - y_k^i$, where $y_k^i = 1$ if the training example belongs to k^{th} class, else $y_k^i = 0$.
3. For the hidden layer, calculate the error by setting $\delta_j^2 = O_j'^2 \sum_i \delta_i^3 W_{ij}^2$, where O_j^2 is output of the excitation function of j^{th} node in the 2^{nd} layer and $O_j'^2$ is the derivative of the same.
4. For the selection layer, calculate the error vector by setting$\delta_j^S = F_j' \sum_i \delta_i^2 W_{ij}^1$, where F_j' is the gradient of the attenuation function in the j^{th} input node.
5. The above W_{ij}^N are the weights between the i^{th} node of the $(N+1)^{th}$ layer and the j^{th} node of the N^{th} layer.
6. For all the m training samples, accumulate the gradient vectors using
$$\Delta_{ij}^1 = \Delta_{ij}^1 + \delta_i^2 O_j^1 \qquad \Delta_{ij}^2 = \Delta_{ij}^2 + \delta_i^3 O_j^2$$
$$\Delta_{jj}^S = \Delta_{jj}^S + \delta_i^S f_j$$
7. The gradient of the neural network cost function is obtained by the following equations ,
$$\frac{\partial J(W,M)}{\partial W_{ij}^l} = D_{ij}^l = \frac{1}{m} * \Delta_{ij}^l, \ for\ j = 0$$
$$\frac{\partial J(W,M)}{\partial W_{ij}^l} = D_{ij}^l = \frac{1}{m} * \Delta_{ij}^l + \frac{\lambda}{m} * W_{ij}^l , \ for\ j \geq 1$$
$$\frac{\partial J(W,M)}{\partial W_{ij}^l} = D_j^S = \frac{1}{m} * \Delta_j^S$$
The λ is the regularization parameter.
8. Update the $\overrightarrow{W^1}$, $\overrightarrow{W^2}$and $\overrightarrow{M}$ using the conjugate gradient descent algorithm [28].
9. After every update, the above steps 1 to 8 are repeated till the target number of iterations are over or the cost function has reached a local minima.

 ISBN: 978-1-61208-250-9

The correctness of the feature selection method in batch mode is verified by using the IRIS dataset [32]. This dataset has four features namely, sepal length (f_1), sepal width (f_2), petal length (f_3), and petal width (f_4) of Iris flower. It is observed that the features f_3 and f_4 get higher importance compared to features f_1 and f_2. This is same as reported in [30] and [31]. Hence the design for the feature selection in batch mode for ANN is confirmed to be consistent to previously reported results. This ANN structure is used in our experiment to analyze the proposed features and previously reported features for human identification from Kinect based skeleton data.

V. EXPERIMENTAL RESULTS

The accuracy of recognizing an individual based on skeleton data is evaluated for the proposed features and compared with the features and results reported in earlier works [21], [22]. We have performed the experiments in four stages. Section V.A compares the proposed features with the one given in [21] for 5 subjects with manual gait data, section V.B gives the results for automatic gait detection, section V.C gives the effect of increase in subjects to 10 and comparison with [22] and section V.D gives the results for feature selection using ANN.

The results of performance comparison are given in terms of F-score, which is the harmonic mean of precision and recall and calculated by (10).

$$F_{score} = \frac{2 * precision * recall}{(precision + recall)} \qquad (10)$$

A. Comparison with [21]

Initially, we have captured data for 5 individuals and compared the results with [21] as Adrian et. al have used 4 subjects in their experiment. The total number of half-gait cycles for 5 individuals is approximately 700 with an average of 140 half-gait cycles per individual.

As reported in [21], we perform k-means clustering on the features with k as 5 corresponding to 5 subjects. TABLE I shows the confusion matrix in percentage for the feature points as grouped in different clusters. It can be noted that the overall average accuracy of recognition is 25.2% as against the by chance accuracy of 20%.

TABLE I. CONFUSION MATRIX FOR 5 CLUSTERS

Person	Level A	Level B	Level C	Level D	Level E
1	19	**40**	10	31	0
2	19	20	**25**	31	5
3	23	20	20	**34**	3
4	23	21	21	32	**3**
5	**24**	25	16	31	4

In this experiment the half-gait cycles are manually extracted to keep the experimental environment similar to [21]. The results of the k-means clustering are compared with the ANN classification. Out of the 700 features vectors, 372 vectors are used for training and the remaining are used for testing. In our implementation, we have used ANN with one hidden layer of 25 nodes as shown in Figure 5.

A summary of the results for 5 subjects with manual half-gait cycle extraction is shown in TABLE II. This clearly indicates that if we take all the 32 features (14 new features ($f = \{f_A, f_D\}$) and 18 earlier features ($f_{[21]}$) proposed in [21]), the recognition accuracy is maximum at 86% which is substantially higher compared to by chance accuracy of 20%.

TABLE II. PERFORMANCE (F-SCORE IN %) COMPARISON OF DIFFERENT FEATURES AND CLASSIFIERS FOR 5 SUBJECTS WITH MANUAL GAIT CYCLE DETECTION

K-Means with $f_{[21]}$	ANN with $f_{[21]}$	ANN with $f = \{f_A, f_D\}$	ANN with $f = \{f_{[21]}, f_A, f_D\}$
25.2	34	69	86

B. Automatic Half-Gait detection

The accuracy of the half-gait detection algorithm (section III) is measured by comparing the ground truth generated by the manual gait detection. The precision of the algorithm is 1, recall is 0.79 and the F-score is found to be 0.88. The accuracy of the recognition of an individual for 5 subjects using automatic gait cycle detection is given in TABLE III. It can be seen that the performance has slightly degraded compared to the manual gait detection (TABLE II).

TABLE III. PERFORMANCE (F-SCORE IN %) COMPARISON OF DIFFERENT FEATURES AND CLASSIFIERS FOR 5 SUBJECTS WITH AUTOMATIC GAIT CYCLE DETECTION

ANN with $f_{[21]}$	ANN with $f = \{f_A, f_D\}$	ANN with $f = \{f_{[21]}, f_A, f_D\}$
33	68	64

A sample confusion matrix for 5 subjects is given in Figure 8 where half-gait cycle is detected automatically and only 14 new features $f = \{f_A, f_D\}$ are used.

```
fscoreTest = 0.682067
a    b    c    d    e  <-- classified as
40   11   36   0    1  | a
1    29   8    0    1  | b
33   9    20   0    3  | c
1    0    0    57   0  | d
0    1    1    0    76 | e
```

Figure 8. Confusion matrix for 5 persons using 14 new features

The results in the remaining section of this paper are based on automatic gait detection.

C. Effect on Increase in Subjects

It is important to analyze the effect in accuracy results as we increase the number of subjects. Pries et. al in [22] have used 9 subjects for their experiments where they have found Naive Bayes classifier to perform best on the 14 features, $f_{[22]}$ proposed by them. As the dataset used by [22] is not publicly available, we have generated a larger dataset of 10 subjects with 8 men and 2 women. The comparison of different features and classifiers are shown in TABLE IV. It can be seen $f = \{f_{[21]}, f_A, f_D\}$with ANN classifier performs best having recognition accuracy of 55% which is much better than chance accuracy of 10%.

TABLE IV. PERFORMANCE (F-SCORE IN %) COMPARISON OF DIFFERENT FEATURES AND CLASSIFIERS FOR 10 SUBJECTS

ANN with $f_{[21]}$	ANN with $f = \{f_A, f_D\}$	ANN with $f = \{f_{[21]}, f_A, f_D\}$	Naïve –Bayes with $f_{[22]}$
27	52	55	51.3

A sample confusion matrix on the test data for ANN with $f = \{f_{[21]}, f_A, f_D\}$ is shown in Figure 9. It can be seen that the recognition for the subject "a" is worst affected whereas others are quite reasonable as seen from high values in the diagonal entries.

```
precision = 0.548295, recall = 0.552487, fscoreTest = 0.550383
confusion_mat
a   b   c   d   e   f   g   h   i   j   <-- classfied as
20  8   47  0   5   3   11  0   1   5   |   a
4   49  10  0   8   8   0   0   0   21  |   b
7   11  52  0   2   5   13  0   2   8   |   c
2   0   0   60  0   0   2   10  26  0   |   d
0   0   1   0   92  3   0   0   0   4   |   e
10  10  5   0   2   44  2   0   0   27  |   f
26  3   20  0   1   3   45  0   1   1   |   g
0   0   0   15  0   0   0   82  3   0   |   h
6   0   1   4   0   0   16  0   73  0   |   i
1   22  0   0   13  25  1   0   0   38  |   j
```

Figure 9. Confusion matrix for 10 persons using 32 features

It is to be noted that the recognition reported here is based on every half-gait cycle, however in actual scenario the recognition will be done on a recording of a session where the decision will be taken on multiple gaits. Considering this fact, we can clearly say that the recognition is 90% (9 identified correctly out of 10) where the entry corresponding to the maximum number in the confusion matrix is taken to identify an individual.

D. Feature Selection using ANN

The feature selection method described in section IV.C for batch mode using ANN is applied on the feature vector $f = \{f_{[21]}, f_A, f_D, f_{[22]}\}$ to understand the more important features against the least important. For this experiment, we initialize attenuation function $F(M_i)$ to 0.05. After the ANN training using the feature selection technique, the values of $F(M_i)$ are shown in Figure 10 for all the 46 features. The first 18 features ($f_{[21]}$) are the one reported in [21], the next two features are based on area (f_A), the next 12 features are based on centroid distances (f_D) and the final 14 features ($f_{[22]}$) are as reported in [22]. It can be seen that most of the static features reported in [22] are having much higher importance than the dynamic features. The skeleton data captured using Microsoft SDK is not noise free and as the dynamic feature captures the property of persons walking in time, the uniformity in feature generation could be lost due to noise incorporation during walk, e.g different angle could be different in different gait cycles. On the other hand the static features capture the relative distances between joints and more stable in computation. Given the raw skeleton information using Microsoft SDK and Neural network as a classifier, these static features get more importance in person identification.

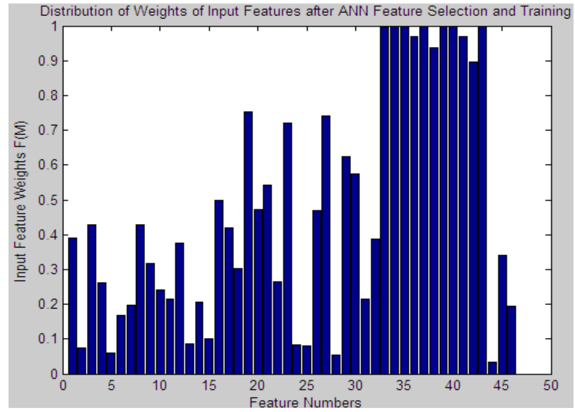


Figure 10. The value of attenuation function F(M) for different features $f = \{f_{[21]}, f_A, f_D, f_{[22]}\}$

The recognition accuracies for 10 subjects with and without feature selection technique are given in TABLE V. It can be seen that with simultaneous feature selection and classification the F-score has increased to 0.62.

TABLE V. PERFORMANCE (F-SCORE IN %) COMPARISON WITH AND WITHOUT FEATURE SELECTION FOR 10 SUBJECTS

ANN without feature selection having features as $f = \{f_{[21]}, f_A, f_D, f_{[22]}\}$	ANN with feature selection having features as $f = \{f_{[21]}, f_A, f_D, f_{[22]}\}$
53	62

VI. CONCLUSION

In this paper we have presented few static and hybrid model based features for gait recognition solution using Microsoft Kinect. The static feature proposed is the area encompassed by the upper body and the hybrid features include the area encompassed by the portion below hip and the distances between the body-centroid and the centroids derived from the joints of the upper and lower limbs. The accuracy of the gait recognition is compared with the methods proposed in [21] and [22]. ANN based connectionist system [30] is used to perform simultaneous feature selection and classification in the batch mode. Results indicate that the static features get higher importance compared to the dynamic features. Moreover, the hybrid features proposed in this paper are having a great importance compared to the dynamic features proposed earlier in the ANN based feature selection process. Future scope of research lies in removing the noise from the skeleton data before extracting dynamic features and also investigates their importance in the feature selection process. Future research is also aimed to use other sources of human identification along with skeleton data from Kinect, to further improve the recognition accuracy. Moreover the effect of using multiple Kinect and camera needs to be analyzed in future.

ACKNOWLEDGMENT

We would like to acknowledge the colleagues of our Cyberphysical Systems Innovation Lab, Tata Consultancy

Services, who have helped us in data capturing activity in spite of their busy schedule in project work.

REFERENCES

[1]. C. Liying, S. Qi, S. Han, C. Yang, Z. Shuying, "Design and implementation of human-robot interactive demonstration system based on Kinect", 24th Chinese Control and Decision Conference (CCDC), May 2012, pp.971-975.

[2]. Arras, K.O., Lau, B., Grzonka, S., Luber, M., Mozos, O.M., Meyer-Delius, D., Burgard, W. "Range-Based People Detection and Tracking for Socially Enabled Service Robots", Prassler, E., et al. (eds.) Towards Service Robots for Everyday Environ. STAR, vol. 76, pp. 235–280. Springer, Heidelberg 2012.

[3]. H. Uwe, H. Sebastian Hommel, B. Michael, and D. Michael, "Face Detection and Person Identification on Mobile Platforms", E. Prassler et al. (Eds.): Towards Service Robots for Everyday Environ. STAR 76, pp. 227–234. Springer, Heidelberg 2012.

[4]. http://www.idiap.ch/scientific-research/themes/biometric-person-recognition(last accessed on 9th Oct 2012)

[5]. X. Qinghan, "Technology review - Biometrics-Technology, Application, Challenge, and Computational Intelligence Solutions", IEEE Computational Intelligence Magazine 2007, vol.2, pp. 5-25.

[6]. D. Kim and J. Paik, "Gait recognition using active shape model and motion prediction", *IET Computer Vision 2010,* vol. 4, pp. 25-36.

[7]. N. Borghese, L. Bianchi, and F. Lacquaniti, "Kinematic determinants of human locomotion" J. Physiology 1996,494(3):863.

[8]. A. K. Jain, R. Bolle, and S. Pankanti, Biometrics: personal identication in networked society, Kluwer Academic Publishers 1999.

[9]. Amit Kale, Aravind Sundaresan, A. N. Rajagopalan, Naresh P. Cuntoor, Amit K. Roy-Chowdhury, VolkerKrügerand Rama Chellappa, "Identification of Humans Using Gait", IEEE Transactions On Image Processing September 2004, vol. 13, no. 9.

[10]. R. Zhang, C. Vogler, and D. Metaxas, "Human gait recognition", IEEE Comp. Soc. Conf. On Computer Vision and Pattern Recognition Workshops 2004, pages 18–18.

[11]. Wang, Jin, She, Mary, Nahavandi, Saeid and Kouzani, Abbas, "A review of vision-based gait recognition methods for human identification", DICTA 2010: Proceedings of the Digital Image Computing: Techniques and Application, IEEE, Piscataway, N.J., pp. 320- 327.

[12]. C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification", Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition 2002, pp. 372-377.

[13]. C. Yam, M. S. Nixon and J. N. Carter, "Automated person recognition by walking and running via model-based approaches", Pattern Recognition 2004, 37:1057-1072.

[14]. S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanIDgait challenge problem: data sets, performance and analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence 2005, vol. 27, pp. 162-177.

[15]. J. Han and B. Bhanu, 2, "Individual recognition using gait energy image", IEEE Transactions on Pattern Analysis and Machine Intelligence 2006, vol. 28, pp. 316-322.

[16]. A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates", IEEE Transactions on Pattern Analysis and Machine Intelligence 2001, vol. 23, pp. 257-267.

[17]. L. Jianyi and Z. Nanning, 2007, "Gait History Image: A Novel Temporal Template for Gait Recognition", IEEE International Conference on Multimedia and Expo, 2007, pp. 663-666.

[18]. C. Chen, J. Liang, H. Zhao, H. Hu, and J. Tian,2009, "Frame difference energy image for gait recognition with incomplete silhouettes", Pattern Recognition Letters, vol. 30, pp. 977-984.

[19]. Z. Xue, D. Ming, W. Song, B. Wan, and S. Jin, "Infrared gait recognition based on wavelet transform and support vector machine," Pattern Recognition 2010, vol. 43, pp. 2904-2910.

[20]. W. Liang, T. Tieniu, N. Huazhong, and H. Weiming, "Silhouette analysis-based gait recognition for human identification," IEEE Transactions on Pattern Analysis and Machine Intelligence 2003, vol. 25, pp. 1505-1518.

[21]. Adrian Ball, David Rye, Fabio Ramos and Mari Velonaki, "Unsupervised Clustering of People from 'Skeleton' Data", HRI 2012, Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, 2012.

[22]. Preis J., Kessel M., Werner M., Linnhoff-Popien C,"Gait Recognition with Kinect", 1stInternational Workshop on Kinect in Pervasive Computing, Newcastle, 2012.

[23]. S. Aravind, R. Amit, and C. Rama, "A hidden Markov model based framework for recognition of humans from gait sequences", Proceedings of the IEEE International Conference on Image Processing, 2003, pp. II-93-6 vol.3.

[24]. C. BenAbdelkader, R. Cutler, and L. Davis, "Motion-based recognition of people in EigenGait space", Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002, pp. 267-272.

[25]. Y. Mu and D. Tao, "Biologically inspired feature manifold for gait, recognition", Neurocomputing 2010, vol. 73, pp. 895-902.

[26]. R. Hecht-Nielson, Neurocomputing, Addison-Wesley: New York, Chapter 5, 1990.

[27]. D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning internal representations by error propagation", Institute for Cognitive Science Report 8506, San Diego: University of California, 1985.

[28]. W. W. Hager and H. Zhang, "A new conjugate gradient method with guaranteed descent and an efficient line search", SIAM J. Optim. 2005, 16, pages 170-192.

[29]. M. Hagiwara, "A simple and effective method for removal of hidden units and weights", Neurocomputing, vol. 6, pp. 207-218, 1994.

[30]. N. R. Pal and K. K. Chintalapudi, "A connectionist system for feature selection", Neural Parallel Sci. Comput. 1997, vol. 5, no. 3, pp. 359–381.

[31]. Chakraborty, D., Pal, N.R.: Selecting Useful Groups of Features in a Connectionist Framework., IEEE Transactions on Neural Network (2008), pp. 381-396

[32]. E. Anderson, "The irises of the Gaspe Peninsula," Bull. Amer. IRIS Soc., vol. 59, pp. 2–5, 1935

Robot Learning Rules of Games by Extraction of Intrinsic Properties

Grégoire POINTEAU, Maxime PETIT, Peter Ford DOMINEY
Robot Cognition Laboratory
INSERM Stem Cell and Brain Research Institute
Lyon, France
gregoire.pointeau@inserm.fr, maxime.petit@inserm.fr, peter.dominey@inserm.fr

***Abstract*—A major open problem in human-robot interaction remains: how can robots learn from non-technical humans? Such learning requires that the robot can observe behavior and extract the sine qua non conditions for when particular actions can be produced. The observed behavior can be either the robots own explorative behavior, or the behavior of humans that it observes. In either case, the only additional information should be from the human, stating whether the observed behavior is legal or not. Such learning may mimic the way that infants learn, through interaction with their caregivers. In the current research we implement a learning capability based on these principals of extracting rules from observed behavior using ”Human-Robot” interaction or ”Human-Human” interaction. We test the system using three games: In the first, the robot must copy a pattern formed by the human; in the second the robot must perform the mirror action of the human. In the third game, the robot must learn the legal moves of Tic Tac Toe. Interestingly, while the robot can learn these rules, it does not necessarily learn the rules of strategy, which likely require additional learning mechanisms.**

***Index Terms*—learning machine; robotics; iCub; *Reactable*; human-robot interaction; human feedback**

I. Introduction

How is a robot able to understand a new game just by watching it or by playing it without any prior assumption ? How can he differentiate two different games without explanation only by the simplest feedback : a simple ”yes” or ”no” from a teacher ? What is the limit of simple imitation [1] ? The answer lies in the understanding of the actions [2]: the intrinsic properties of each move. And from this properties, the extraction of *sine qua none* conditions. In our studies, we want to teach a robot to play different board games (as a Tic-Tac-Toe for example) without giving it any assumption about the game. The idea of learning a new concept as it has been done in Wu in [3] but here, we will not use any Bayesian statistics but only simple deterministic algorithm. Breazeal et al. have shown the efficiency of socially guided exploration ([4] and [5]) and social interaction also has been shown crucial in learning ([6] and [7]). Most board games use a set of ”natural” rules as the turn taking, that we give to the robot. In our case, we use an humanoid robot called *iCub*, in interaction with an Human through an interactive table [8]: the *Reactable* [9]. Only with a feedback : *”yes or no”* from the user, for each move (is the move is legal or not), the robot is able to discriminate different games and to learn a new game. Also, the robot can learn a game, just by watching 2 humans playing. In this paper we will first see the global architecture and overview of the system, then the experimental protocol and the games used, then the learning part directly, and finally the results of simulated ”Human-Robot” interactions and ”Human-Human” interactions.

II. System Design Overview

A. Robotic Platform : iCub

The iCub [10] is a humanoid robot open-source platform. It has 53 actuated degrees of freedom (with 19 for each 5-fingers hand) and with a height of 104 cm, has a morphology approximating that of a 3 year-old child.

 ISBN: 978-1-61208-250-9

The distributed modules used to run the robot are interconnected with the open source library YARP [11] through ”ports”. The ports can exchange data over different network protocols such as TCP and UDP.

The motor control is managed by a Passive Motion Paradigm approach [12], where hand trajectories for reaching actions are computed according to virtual force fields with attractor (target) and repellor (obstacles).

B. Sensor System : Reactable

In the current research we extend the perceptual capabilities of the iCub with the *Reactable*. This is a tabletop tangible interface [9] licensed by *Reactable* Systems. This allows the human and the iCub to manipulate objects in a shared space that the iCub can perceive with high precision. An infra-red illumination beneath the table allows reacTIVision [13], a detection system based on an infrared camera, to accurately and in real-time identify and track tagged object (using fiducial markers [14]) placed on the translucid table. It can also recognize fingertips (cursor) which allow the user to manipulate the digital information (music, game, ...) with real-world objects or direct tactile control (see Figure. 1).

C. Knowledge Base : OPC

The different knowledge of the robot is centralized in a database called OPC for ObjectPropertiesCollector, and grouped accordingly into some enitites : the informations provide in particular by the *Reactable* (e.g. position x and y of an object with the *Reactable* id i) are merged with some ground knowledge (e.g. the object with *Reactable* id x is called ”$cross$” and has a size of x,y,z) or reasoning conclusion (e.g. ”$onTable1$” if an object is placed on the table) in order to have a full picture about this object. These entities can be accessed with unique identifiers and are managed dynamically in real-time.

The robot knowledge is initialized through the OPC set-up. The database includes :

- locations : 3x3 squares (from A-1 to C-3) to obtain a board game setup, and two other places to put respectively the cross and circle object when unused.
- objects : Two kind of pawns, cross in blue and circle in red, and a special object, the eraser to undo a move if needed.

D. Spoken Language and Supervisor : CSLU-RAD

The spoken language interaction is implemented with the CSLU Toolkit [15] Rapid Application Development (RAD) under the form of a finite-state dialogue system where the speech synthesis is done by Festival and the recognition by Sphinx-II. In addition to provide a human-robot language-based interaction during games, it is used in particular to extract oral feedback from the user about the move quality done by the robot.

This state machine is built in what we called the Supervisor. It has two main functions : guide the human into the games, inviting him to act or choose between options, and to control the robot behavior (including speech of the iCub) using human spoken feedback, the knowledge base or some external module, like BoardGameLearning.

E. BoardGameLearning

BoardGameLearning is a module developped in c++, responsible for the statistical analysis and the learning. We will explain how it works in paragraph IV.

III. Experimental Protocols

A. Games used

For our studies, we use 3 different games, with different properties that we wanted to test with only one learning module. All of them work with a turn taking managed by the supervisor (see II-D and Figure 2).

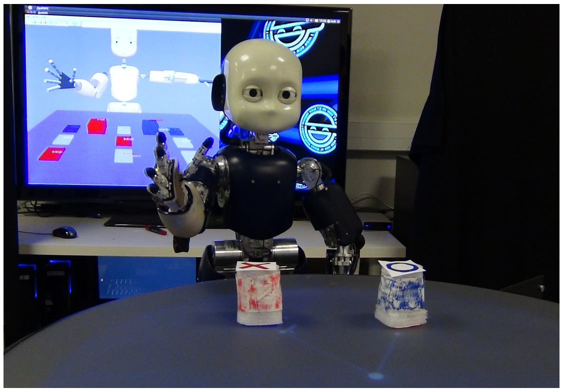


Fig. 1. Normal functioning of the system with iCub and objects on the *Reactable*. A screen behind the iCub displays the iCub's internal representation of the environment including the board game on the *Reactable*. The iCub can manipulate the two stamp for marking of cross or circle in the game. The Reactable detect the position of the objects.

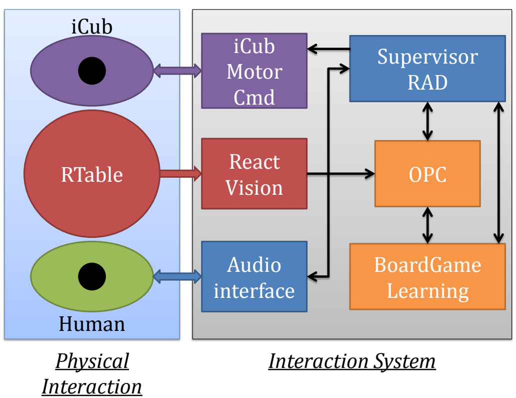


Fig. 2. System architecture. In purple is the robot-related. In blue is the supervisor and interaction related. In orange in the computation and learning module part. In red, the Reactable.

1) Pattern game: The first game is a Pattern Game. The iCub and the Human, with alternating turn taking, have to fill the board with a predefined pattern. For example the pattern in for the first and the third row, to fill with cross. Each game has the same pattern (otherwise, the robot can't learn). For the Pattern Game, the important properties are more related to what was learned in the previous game, rather than the current one.

2) Mirror game: The second game is a Mirror Game. For the Mirror game, the iCub has to play the same move as the previous one by the Human. The Human can play any color he wants, but only on an empty case. For the Mirror game, we had to add one precision for the robot : the *Human and the Robot don't have the same rules*. The Human has to play anywhere one the board, on a free case. The Robot has to reproduce the same move as the Human. This is like a field player and a goal-keeper in football, the game is the same but the rules are not the same for every player. This is a precision that we have to add in the system.

3) Tic-Tac-Toe: The third game is a classic Tic-Tac-Toe (TTT). For the TTT, a legal move is that each player has to play in an empty case, with a different color that the one used by the previous player. For the TTT, the properties involved are mainly the location (free or occupied) and the color and only the current game really matters.

B. *Progress of the game*

The Robot can learn two different ways. The first one is to play with the Human, where the Human is considered as the *teacher*, and the Robot as the *student*. In the second way, the robot can learn only by watching two Humans playing together, with one of them considered as the *teacher*, and the other one, as the *student*.

The teacher always will have the knowledge of the rules of the game, and will give feedback for every move (the student's move and the teacher's move).

IV. How to learn legal moves ?

This part of the module called *boardGame-Learning* (BGL) has the aim to distinguish legal moves from illegal moves. A legal move is a move authorized by the game, but not necessarily a *"good"* move. For example for a game like Tic-Tac-Toe, a illegal move is to put a different color of the on of the previous player, only on a

TABLE I
DEFINITION OF THE DIFFERENT PROPERTIES, WITH THE NUMBER OF POSSIBILITIES FOR EACH (G : NUMBER OF GAMES PLAYED, C : NUMBER OF TURNS PLAYED DURING THE CURRENT GAME)

Properties	Definition	Pos.
Location	Free / Occupied	2
Location	Which turn this spot has been played in the current game	C
Location	How many time this spot has been played in the previous games	G
Piece	How many time this piece has been played in the previous games	G
Piece Location	How many time this piece has been played at this location in the previous games	C
Color	Same as previous Different from previous	2
Color Piece	How many time this piece has been played with this color in the previous games	G
Color Location	How many time this piece has been played at this location in the previous games	G

unoccupied case.

The learning system is based on a feedback from the user. We start with the idea that the robot doesn't know the game at all. He just knows that he has different kind of pieces (for example : "pawn", "bishop"... In the case of Tic-Tac-Toe, he only has "pawn"), and a board in front of him and he has to move a piece of any color on the board. The robot also has the knowledge of turn taking.

A. Random trials

At the beginning of the game, the iCub will explore the board and his possibilities, and he will expect a feedback from the Human if his move is legal or not. With this feedback, the robot will be able to extract pertinent properties of the legal moves. Each time a move is played, the robot will increment his statistics according to the properties of the move. The properties concerned are the location, the color, the kind of piece and interaction between these properties (see Table I) :

B. Probabilistic Exploration

Each turn, the iCub will pick a move, according to the probabilities extracted from his experience. The probabilities of each properties are given by the following formula :

$$S = \frac{\sigma + N_{hits}}{\sigma * P + N_{tries}} \quad (1)$$

where :

- σ is a learning rate constant
- N_{hits} is the number of good moves with this property
- N_{tries} is the number of tries of moves with this property
- S is the score of each property

We can easily see that $N_{hits} \leq N_{tries}$. Also σ is a learning rate which correspond to : *"after how many tries should I be certain of my experience ?"*. A high σ will delay the influence of experience into the decision. The probabilities correspond to the normalization of the scores in the aim to have a sum of the probabilities of every possibility of a property, equal to 1. For the first moves, the iCub will randomly pick a move, but following his*"intuition"*. This *"intuition"* can be improved by the use of threshold set on the probabilities.

In order to improve his exploration, after an illegal move, the iCub will try the same spot but will change the color then, the piece, and if finally the problem comes from the location and not the piece, he will change his location.

C. Sine qua non conditions

After a few tries, the iCub will be able to extract some *sine qua non* conditions. For this, he has to check if for one property : $N_{hits} = 0$; and : $N_{tries} > \theta$, or simply : "After θ moves with a given property, I have always failed, this should be a *sine qua non* condition" not to make this move. And he will set the probability to 0. This assume that after θ tries, all the configuration with one property fixed, have been seen. The iCub will be able to tell with certainty if, according to the

Human a move is legal or not.

The value of θ will be discussed in the paragraph. V-D.

V. Simulation

In order to test the system, we proceed with simulation experiments

A. Games used

As described in paragraph III-A, we used 3 different games for our tests : the Tic-Tac-Toe (TTT), a Mirror Game, and a Pattern Game.

B. Simulation of the teacher moves

For the teacher moves, we coded the rules of each game. We encoded a function to pick a legal move, and a function to attribute a reward to a student's move. In this case, the teacher has no strategy, but it doesn't matter, because, we only focus on the legal moves (see paragraph IV).

C. Results

To test our system, we also simulated the Human part. The simulated Human knows the exact rules of each game, and always picks a legal move for each game. We also made a auditor of legal or illegal move to check. But for each new game, we have to implement the rules of the Human and the auditor.

For 200 simulations, we made the iCub and the simulated Human to play together and the results are summarized in the Fig. 3, Fig. 4 and 5.

1) Result Pattern game: As shown on the Fig. 5, the probability of succes with our module increase rapidly, to reach 100%. However, we can see some differences between different values of the parameter θ. We will discuss about this parameter in paragraph V-D.

2) Result Mirror game: For the mirror game, we can see that the curve of the random pick, seems to stabilize at 0.05, which correspond to $\frac{1}{2}$ for the color, multiplied by $\frac{1}{9}$ for the location, i.e. : $\frac{1}{18} \simeq 0.0556$.

The random curves correspond to a random pick from the *iCub* (18 possibility : 9 case and 2 colors).

3) Result Tic-Tac-Toe: For the Fig. 3 we can see that a random pick we be more successful for the first pick. It is just an effect of the game. Indeed, the Human always started to play. After one move, 8 cases were free, and the iCub had a probability of $\frac{1}{2}$ to take the good color, and of $\frac{8}{9}$ to choose a good spot i.e. : $\frac{4}{9} \simeq 0.444$. But this is only for the first move, because the situation changes in each simulation according to the success or not of the robot. This is why for the first random pick, we have a good probability of success.

D. Influence of the θ parameter.

As we explain earlier, θ is a constant that we can see as a time constant. It corresponds to: "when do I assume that I have seen enough moves to apply a threshold on my predictions ?". As we can see in Fig. 3, for the TTT, as soon as this θ is reached, a rate of succes of 100% is also reached. This mean that after 10 moves, the iCub has seen enough to understand. On contrary, for the Pattern game, the results are better for a higher θ (Fig. 5). After only 10 moves, the iCub *"thinks"* that he has seen enough to be sure of his decision, but the game is too complicated to be understood after so few moves. The whole space of possibles is not yet totally explored.

This parameter can be set according to the complexity of the game.

VI. Human Interaction

A. Experiment Description

This experiment involves the iCub and the Human physically interacting, playing a game, using objects on the *Reactable* as pieces (see Fig. 7).

The robot knowledge is initialized through the OPC setup. The database includes :

- 11 locations : 9 brown squares (3x3) from A-1 to C-3 and 2 squares for the objects origin place, forming the game "board".

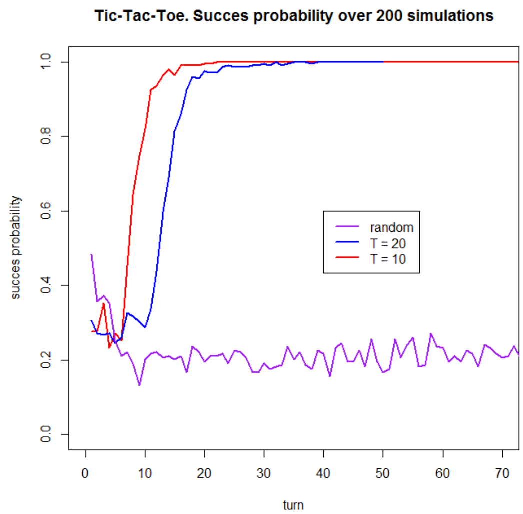


Fig. 3. Results of the learning module on a Tic-Tac-Toe. X-axis is the number of turns passed, Y-axe is the rate of success of the X^{th} move over 70 simulations. The blue curves is for $T = 20$, red for $T = 10$ and the purple curve is in the case of a totally random pick. T is the delay before checking the *sine qua non* conditions.

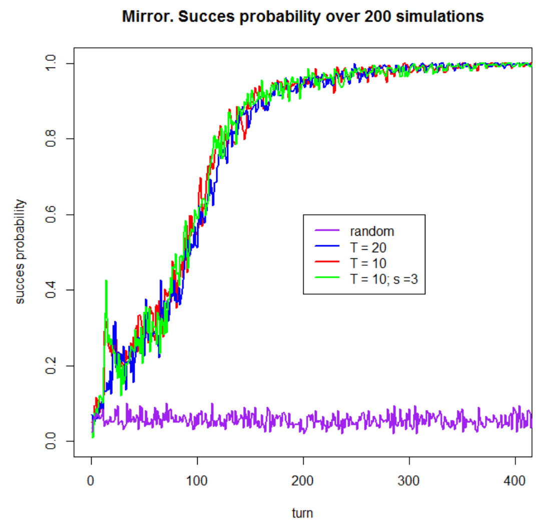


Fig. 4. Results of the learning module on Mirror game. X-axe is the number of turns passed, Y-axis is the rate of success of the X^{th} move over 200 simulations. The blue curves is for $T = 20$, red and green for $T = 10$ and the purple curve is in the case of a totally random pick. T is the delay before checking the *sine qua non* conditions. All the curves are set with $\sigma = 6$ except the green curve for the mirror game where $\sigma = 3$.

- 3 objects : a cross (red) and a circle (blue), which are the played pawn and will change the color of the squares where there are. An eraser is also available to do some correction (i.e. change the color of the square) on the board in case of mistakes.

The objects are like stamps, or playing pieces to be put on the board. To take a turn, the player need to put the object he wants to use from the origin place to the location wanted. An algorithm check where a location is intersected with an object and will change his color accordingly. The user, in case of human, has to say it has moved a pawn (”Done”) in order to detect what is the move played by the user (What pawn has been placed where?). It is automatically launched when the iCub is playing at the end of his move. Next the user has to bring back the piece to its corresponding origin location and invite the other player to play when it is done. If the iCub has played, and if it is in a learning mode, the robot asks the human to give a feedback about the move (illegal, bad, good or draw/win at the end of a game). If the move is illegal, the iCub will then play again until he find a proper one.

B. Generality and perspectives

The iCub can also learn just by watching 2 Humans playing, but he will not be able to extract some *sine qua non* conditions, if the Humans always play some legal moves (N_{hits} will always be N_{tries}, except if one of the Humans is a beginner too).

We have seen that we give to the iCub 2 presuppositions : the turn taking, and if the 2 players have the same rules (mirror game) or not. The concept of turn taking, as it is learned in Broz et al.[16] might be interesting to include in the robot. But this is a very primitive instinct and it might be much more interesting for us, to learn the different roles of each player.

 ISBN: 978-1-61208-250-9

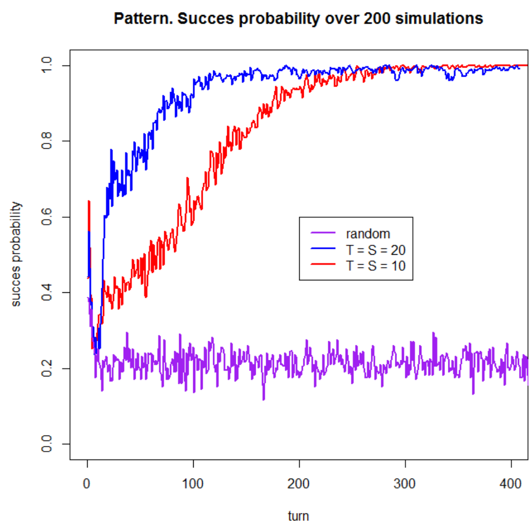


Fig. 5. Results of the learning module on Pattern game. The pattern to find was three crosses on the left and right column, and three circles in the middle column. X-axe is the number of turns passed, Y-axis is the rate of success of the X^{th} move over 200 simulations. The blue curves is for $T = \sigma = 20$, red and green for $T = \sigma = 10$ and the purple curve is in the case of a totally random pick. T is the delay before checking the *sine qua non* conditions.

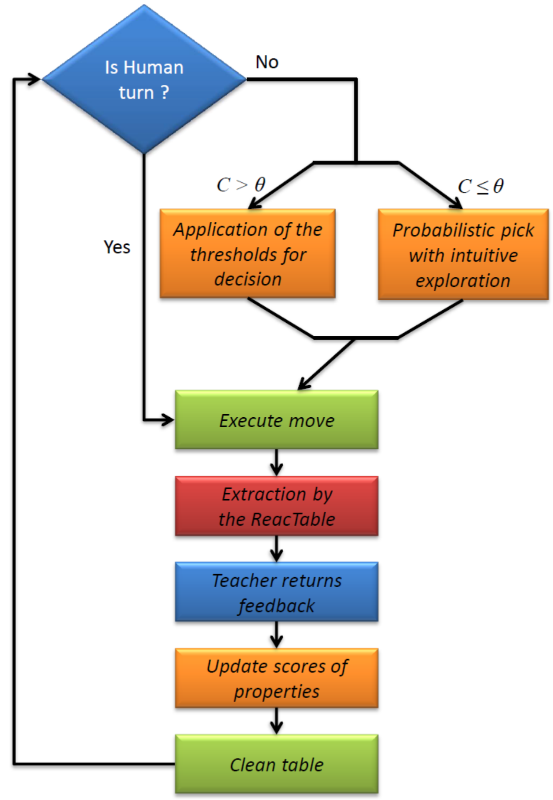


Fig. 6. Progression of moves with our learning system. C is the number of move since the beginning of the learning. θ is the confidence threshold. It can have one or two humans playing. One of them is always the teacher and will give the feedback for each turn (legal-illegal. Win-lose.). The Human/Teacher actions are in blue. In orange is the computation part of the system. In red is the acquisition with the *Reactable* of each move. In green is the player (Robot or Human) action : play and clean. The "execute action" can be done by the three kind of player : Teacher, Human and Robot.

VII. Conclusion and Futur Work

We have been able to manage the learning of different board game for our robot, simply by using the experience, and intrinsic properties of action. But a problem remains in the learning of "good" or "bad" moves. For this we are currently developing a system based on experience and on the outcome of each game to learn the winning combination and not to be tempted to reproduce a failing behaviour.

In our case, the tasks that the system can be used to solve were games, but we can imagine to learn cooperation as is the case for the Pattern game. The long term aim of this system, is not only to learn board game, but to understand more complex concepts. For example, with the Pattern Game, we can teach some concepts as : *"a row", "a line", "left", "right", "before", "after"*... All concept, put together, can reach to emergence of more complex rules, or understanding of new concepts.

Acknowledgment

This research was supported by the EFAA Project, funded by FP7-ICT- Challenge 2 Cognitive Systems, Interaction, Robotics Grant Agreement no: 270490

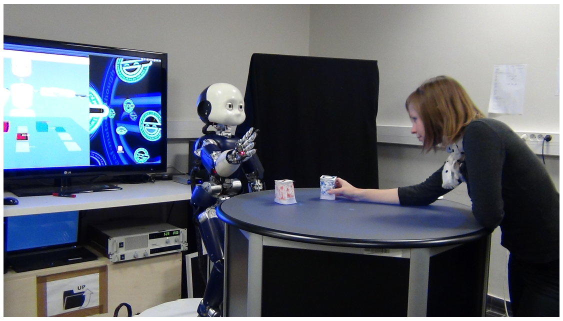


Fig. 7. Normal working of the module with a subject and the Reactable.

References

[1] C. Nehaniv and K. Dautenhahn, "Of Hummingbirds and Helicopters: An Algebraic Framework for Interdisciplinary Studies of Imitation and Its Applications." *Interdisciplinary Approaches to Robot Learning*, vol. 24, 2000.

[2] M. Petit, S. Lallee, J. Boucher, G. Pointeau, P. Cheminade, D. Ognibene, E. Chinellato, U. Pattacini, I. Gori, U. Martinez-Hernandez, H. Barron-Gonzalez, M. Inderbitzin, A. Luvizotto, V. Vouloutsi, Y. Demiris, G. Metta, and P. Dominey, "The Coordinating Role of Language in Real-Time Multi-Modal Learning of Cooperative Tasks," *IEEE TAMD*, pp. 3539–3544, 2012.

[3] X. Wu and J. Kofman, "Human-Inspired Robot Task Learning from Human Teaching," *2008 IEEE International Conference on Robotics and Automation Pasadena, CA, USA*, 2008.

[4] C. Breazeal and A. Thomaz, "Learning from Human Teachers with Socially Guided Exploration," *2008 IEEE International Conference on Robotics and Automation*, pp. 3539–3544, 2008.

[5] D. Grollman and O. Jenkins, "Dogged Learning for Robots," *2007 IEEE International Conference on Robotics and Automation, Roma, Italy*, 2007.

[6] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot." *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 37, 2007.

[7] N. Mirza, C. Nehaviv, K. Dautenhahn, and R. Boekhorst, "Developing Social Action Capabilities in a Humanoid Robot using an Interaction History Architecture," *Proc. IEEE-RAS Humanoids 2008*, 2008.

[8] S. Lalle, U. Pattacini, J. Boucher, S. Lemaignan, A. Lenz, C. Melhuish, L. Natale, S. Skachek, K. Hamann, J. Steinwender, E. Sisbot, G. Metta, R. Alami, M. Warnier, J. Guitton, F. Warneken, and P. Dominey, "Towards a Platform-Independent Cooperative Human-Robot Interaction System: II. Perception, Execution and Imitation of Goal Directed Actions Proceedings," *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.

[9] G. Geiger, N. Alber, S. Jordà, and M. Alonso, "The Reactable: A Collaborative Musical Instrument for Playing and Understanding Music," *Her&Mus. Heritage & Museography*, no. 4, pp. 36 – 43, 2010.

[10] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori, *The iCub humanoid robot: an open platform for research in embodied cognition*. New York, New York, USA: ACM Press, Aug. 2008. [Online]. Available: http://portal.acm.org/citation.cfm?id=1774674.1774683

[11] G. Metta, P. Fitzpatrick, and L. Natale, "YARP: Yet Another Robot Platform," *International Journal of Advanced Robotic Systems*, vol. 3, no. 1, 2006. [Online]. Available: http://eris.liralab.it/yarp/

[12] V. Mohan, P. Morasso, G. Metta, and G. Sandini, "A biomimetic, force-field based computational model for motion planning and bimanual coordination in humanoid robots," *Autonomous Robots*, vol. 27, pp. 291–307, 2009.

[13] M. Kaltenbrunner and R. Bencina, "reacTIVision: a computer-vision framework for table-based tangible interaction," in *Proceedings of the 1st international conference on Tangible and embedded interaction - TEI '07*. New York, New York, USA: ACM Press, Feb. 2007, p. 69.

[14] R. Bencina and M. Kaltenbrunner, "The Design and Evolution of Fiducials for the reacTIVision System," in *3rd international conference on generative systems in the electronic arts*, 2005.

[15] S. Sutton, "Predicting and explaining intentions and behavior: How well are we doing?" *Journal of Applied Social Psychology*, vol. 28, pp. 1317–1338, 1998.

[16] F. Broz, H. Kose-bagci, C. L. Nehaniv, K. Dautenhahn, and A. V. Attention, "Learning behavior for a social interaction game with a childlike humanoid robot."

”Where is Your Nose?” - Developing Body Awareness Skills Among Children With Autism Using a Humanoid Robot

Sandra Costa
Algoritmi Centre
University of Minho
Guimares, Portugal
Email: scosta@dei.uminho.pt

Hagen Lehmann,
Ben Robins
and Kerstin Dautenhahn
Adaptive Systems Research Group
University of Hertfordshire
Hatfield, UK

Filomena Soares
Algoritmi Centre
University of Minho
Guimares, Portugal

***Abstract*—This article describes an exploratory study in which children with autism interact with KASPAR, a humanoid robot, equipped with tactile sensors able to distinguish a gentle from a harsh touch, and to respond accordingly. The study investigated a novel scenario for robot-assisted play, namely to increase body awareness with tasks that taught the children about the identification of human body parts. Based on our analysis of the childrens behaviours while interacting with KASPAR, our results show that the children started looking for a longer period of time to the experimenter, and a lot of interest in touching the robot was observed. They also show that the robot can be considered as a tool for prolonging the attention span of the children, being a social mediator during the interaction between the child and the experimenter. The results are primarily based on the analysis of video data of the interaction. Overall, this first study into teaching children with autism about body parts using a humanoid robot highlighted issues of scenario development, data collection and data analysis that will inform future studies.**

Keywords: *Assistive Technologies; Socially Assistive Robots; Human-Robot Interaction; Body Awareness.*

I. Introduction

Three critical factors for the healthy physical and psychological child’s development are touch, movement and interaction with other humans. Touch is one of the earliest senses developed in human embryos and the most developed sense at birth [1]. Thus, touch plays a key role in the physical, emotional, and psycho-social development. Touch deprivation early in life leads to severe consequences, like complete emotional isolation or lack of trust in others [1]–[3]. Children need up to four hours per day of physical play to accomplish satisfactory sensory stimulation for their proprioceptive and tactile systems in order to develop normally [4]. On one hand, touch can convey affectioned feelings, on the other, it can express pain or discomfort.

Touch can be divided into cutaneous, kinesthetic, and haptic systems [5]. The cutaneous system is constituted of mechanoreceptors set in the skin. This system composes the tactile sense, processing stimulations on the skin. The kinesthetic system is constituted of receptors situated in the muscles, tendons, and joints. The kinesthetic sense allows humans to identify positions and movements of upper and lower limbs and muscle tensions. The haptic sensory system concerns both cutaneous and kinesthetic receptors, but it is associated with an active procedure [5]. Children learn early on to understand and to identify different types of physical contact in order to communicate with other children and adults, building trust relationships, based on the exchange of support and mutual confidence, developing their social relationships.

According to Piaget infants develop object permanence through touching and handling objects [6]. Object permanence is the understanding that objects continue to exist even when they cannot be seen, heard, or touched [7]. Caregivers typically offer organized environments where children can explore, touch and manipulate different materials and where they can be able to ask questions, use their creativity and learn new concepts. Children have to build their own learning experience, with the focus on the reasoning processes [8].

In this study, we used a humanoid, minimally expressive child-sized robot with a static body - KASPAR ([9] for technical details), able to move its arms and head in order to simulate gestures in social interaction. KASPAR has simplified and minimalistic human-like features. The robot’s behavioural repertoire includes expressive postures, it can approximate the appearance and movements of a human without trying to create an ultra-realistic appearance. KASPAR is equipped with tactile sensors that allow it to automatically respond to a gentle or rough touch from the child. The tasks used this study aim to teach the children to identify their body parts, and increase their body awareness. As we can see in the Section II, robots have already been used with children with autism to develop their social and communicative skills with encouraging results. In the present study, the robot is going to be used as a mediator between the child and the experimenter but also as a tool of teaching. Our main research interest is to understand if the robot can help to elicit interactions between an autistic child and another person, and whether it can facilitate the ability to acquire knowledge about human body parts. We want to verify if the robot can help children with ASD (Autism Spectrum Disorder) to learn appropriate physical social engagement. The experiments consisted of 7 sessions with 8 children diagnosed with autism, using qualitative and quantitative measures to evaluate the triadic interaction between the children, the robot, and the experimenter. In this article, we present the analysis of the observations of the first and the last session with the 8 children. To our knowledge this is the first article that studies how to use robots in order to teach children with autism about body parts. Due to the novelty of the subject a main purpose of the article was to develop scenarios, means of data collection and to learn how to analyse the data. This paper is organized as follows. In Sect. 2 will be presented research projets that also use tactile human-robot interactions. Sect. 3 features the procedures during the experiments. The results and the discussion are described in Sect. 4 and 5. Sect. 6 provides the conclusions and future work.

II. Tactile Interaction

As mentioned earlier, touch plays a vital role in human-human interaction. Since it is our goal to transmit extra information to the robot, so it can react predictable and convincing to a human tactile interaction, the robot needs to be equipped with tactile sensing capabilities. The robot's behaviour must appear natural, in order to generate enjoyable interactions.

There are several research projects concerned with the physical contact between humans and robots, presenting various types of sensors to detect these interactions, for example the cheap and robust sensors that can measure force or pressure, changing its resistance, called force-sensing resistors (FSR) [10]. The information provided by tactile sensors aims to increase quality and interpretation of sensor data. We can measure improvements in tactile sensing according to data quality, assessed relatively to detection sensitivity, noise and physical toughness and also its signal interpretation, assessed relatively to computational cost and measurement accuracy [10]. The detected contact should be used to produce compliant robotic behaviours.

Robots for human-robot interactions (HRI) within the current tactile HRI literature can have different shapes [10]. The baby seal Paro [11], the teddy bear Huggable [12], the robotic cat NeCoRo [13], and the child-sized robot KASPAR [9] are some examples of different artificial pets and humanoid robots designed to engage people based upon relational touch interactions. This kind of affective interaction is a growing area of research, especially concerning the target group of people with special needs.

Tactile data contributes to the determination of the Paro's internal state, driving the choice and implementation of a limited number of hand-coded behaviours, similar to those of a real seal [14]. Huggable, the robot teddy bear able to orient itself towards the direction of the human touch through motion in its neck and shoulders. A soft multi-modal sensory skin plus the fur covering its entire teddy-bear-shaped body are able to classify multiple human touch types and perform tactile interactive behaviours [12]. The robotic cat NeCoRo is used to analyse personrobot communication, responding to human voice, movements, and touch. Its multiple sensors, together with artificial intelligence technology produce a real-life-looking robotic cat capable of a playful and natural communication with humans [13]. KASPAR is a robot that has been used e.g. in call-and-response games, where its goal was often to imitate the human partner [9]. In the ROBOSKIN project, researchers develop a robotic skin to provide tactile feedback which was added to KASPAR. One goal is to improve human-robot interaction capabilities in the application domain of robot-assisted play [15].

III. Procedures

The experiments were performed in four different phases: familiarisation, pre-test, practice, and post-test (fig. 1).

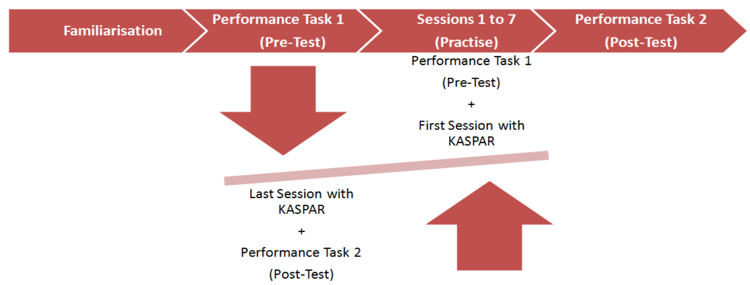


Figure 1. Phases of the study, in this paper only the pre- and post-test, and the analysis of the first and last session of the practice phase will be presented

Before starting the experiments with the robot, the children and the experimenter took part in a familiarisation phase. The goal of this phase was to get acquainted with the children and to integrate the experimenter in the school environment. The experimenter spent one day at the school, in the classroom where the children normally do their activities. A pre-test served as baseline to be compared with the results of a post-test. The task tested in these two phases was the performance task. The post-test phase had the same conditions of the pre-test phase in order to evaluate if the children were able to improve the performance of the task done in the pre-test. In the practice phase, three different activities were introduced according to the children's accomplishment. The next sub-section presents the different tasks associated with each phase.

Each session with the robot was introduced with a Picture Exchange Communication System (PECS) card, that the children usually use in their daily routine to start new activities. The PECS card used for this experiment depicted the KASPAR robot (Figure 2).

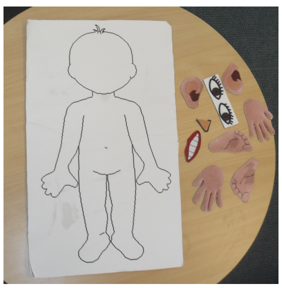
A. *Tasks*

In the performance task, carried out in the pre- and post-test, the children were asked to choose the right place for the different body parts, and place them on a drawing of a little human figure printed on a cardboard (Figure 3). The

 ISBN: 978-1-61208-250-9

Figure 2. PECS card of KASPAR

performance task used in the pre- and post-test took the TEACCH program [16] already used in the classroom by the teachers into consideration.



(a) Beginning of the Task

(b) Task Accomplished

Figure 3. Performance task done in the pre- and post-test

In the practice phase, with 7 sessions of approximately 10 minutes each, there were three different activities, focusing on body awareness skills. The complexity of the activities was different, so whenever the children managed to accomplish the first activity, in the next session they would perform additionally the next more complex activity. If a child did not manage to progress, more sessions were done with the basic activity. The evaluation of the right transition moment to the next level for each child was done by the experimenter based on the opinion of the teachers. At this stage the robot's response was triggered remotely by the experimenter.

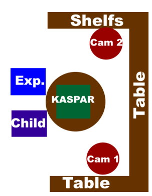
- Activity A: The robot identified one part of its body saying: "This is my head". Then, it asked: "Can you please show me your head?". If the answer of the child was correct, the robot responded with a positive reinforcement like "That's right!" or "Well Done!". If the answer was not correct, the robot encouraged the child to try again, e. g. "Almost. Try again!". The human body parts to be identified were: head, tummy, nose, ears, eyes, hands toes, and mouth.
- Activity B: The robot identified a sequence of human body parts on its own body. For example: head and tummy. Next, it asked the child to point at the same body parts and in the same sequence on her own body. Then, the following step was to use three body parts (e.g., head, tummy and toes). The same type of reinforcement of Activity A was used.
- Activity C: This activity involved the learning from the previous activities together with joint attention and interaction with the experimenter. The robot asked the child to sing together a song about human body parts, and the experimenter encouraged the child to do the gestures that accompanied the song. If the child did not have verbal communication, he was asked to do the same gestures of experimenter (moving their body parts according to the song). The song was chosen based on simplicity and the practical learning approach is normally used in the school to teach other contents.

B. Participants

The eight participants in the experiments were boys with ASD aged six to ten years old. Four of the participants were high-functioning (Group A), the others were low-functioning (Group B), according to the diagnosis of the children. The experimenter was in the room to introduce the robot, and to intervene in case of difficulties. She was also involved in the activity as a facilitator of the interaction, providing guidance and ensuring that the children did not become agitated or damage the robot during the activity. A signed informed consent form was obtained from the parents of each child. This work was granted ethics approval by the University of Hertfordshire.

C. Settings

The experiment took place in a familiar room in the school often used by the children for their activities (Figure 4). The robot was connected to a laptop and placed on a table in the centre of the room. The children were sitting or standing facing the robot.



(a) Room Setup Schematic

(b) Positioning of the participants in the room

Figure 4. Room Setup

The two cameras were placed in such a way that one recorded the face of the child and the other the experimenter during the experiments.

D. Touch Feedback

The robot was equipped with 8 FSR sensors positioned on the right head, left head, right shoulder, left shoulder, right wrist, left wrist, right hand, left hand, right foot, and left foot of the robot. These FSRs only distinguished a gentle from a harsh touch. If the child touched the robot, activating the sensor below the threshold limit, it answered a sentence such as "You are so gentle. Thank you.". If the child touched the robot and activated the sensor above the threshold limit, it answered with a sentence such as "Ouch, you are hurting me.".

The threshold limit was defined in experimental pre-tests. The goal of this feedback is to automatically produce a response to the children's tactile interaction, teaching appropriate physical social engagement, reinforcing suitable behaviours when using touch to interact with another agent.

E. Evaluation Tools

As a qualitative measure, we used a structured interview. As quantitative measures, we used questionnaires, a behavioural analysis, where the children's behaviours were identified and coded with video analysis, and the comparison between the pre- and post-test.

1) Questionnaires: The questionnaire aimed to measure the development of children assessing their skills regarding tactile interaction. The questionnaire was delivered at two different points during the trials. First, before the trials with the children to establish a data baseline for each child. Then, the last evaluation was done at the end of the study, to evaluate the changes in the behaviours of the children. The items were with a 5 point Likert-scale. Three teachers completed the corresponding questionnaires for the children. For each question, space was available for comments, providing information not covered by the response categories. The questions were mainly related to tactile interaction and the knowledge about body parts, such as, 'Does the child use his/her hands to explore novel/unknown objects?' or 'Can the child point or identify parts of his/her body in any way?'.

2) Structured Interview: The structured interview was done with one of the teachers, showing her extracts of the videos of each child. In this interview, we were interested in the perspective of the teacher on the children's behaviours. Mainly, we wanted to know, how the teacher would describe the reactions of the children towards the robot and what usual or unusual behaviours the children showed in the video. Also, the children's social behaviour seen in the video was compared with the behaviour of the children towards teachers and other children in the classroom (tactile interaction, eye gaze, playing with others, among others). After discussing this, the main differences in the children's behaviour in the two videos were discussed, as well as whether the robot could have had an influence on the specific behaviours performed by the children. The interviewed teacher knew only four of the eight children very well and thus only commented on these. Despite this fact, we considered her comments very relevant and included them in this article.

3) Behavioural Analysis: The sessions were examined via video analysis (using the Observer XT 11 program by Noldus). The behaviours coded were the following: looking, touching, following, pointing, imitation, prompts, and identifying body parts. For each coded behaviour (except for looking) the coders marked whether the child showed the behaviour spontaneously or whether the behaviour was prompted by the experimenter. If the child was, for example, touching the experimenter for no specific reason, the behaviour was classified as spontaneous. If the child touched the experimenter after the experimenter said "Where is my nose?", the behaviour was classified as prompted. A behaviour ended if the child stopped exhibiting that behaviour or showed another behaviour, directly related (for example, looking at KASPAR/looking at the experimenter). When the child exhibited behaviours that are not specified in our list, they were not coded. For eye contact turning away ended the behaviour. Turning back immediately and making eye contact again counted as a new behaviour. To ensure inter-rater reliability 10% of the videos were re-coded by a second independent coder (Cohen's kappa k = .63).

4) Comparison between pre- and post-test: When comparing the pre- with the post-test, special attention was paid to the time taken to accomplish the performance task. Some of the children needed help to finish the task, but this help was only provided when it was verified that the children were not able to solve the performance task.

IV. Results

The collected data from the questionnaires, the behavioural analysis, and the comparison between pre- and post-test were statistically analysed, and a descriptive evaluation was made based on the structured interview.

A. Questionnaires

To determine how the responses of the teachers on the written questionnaires matched for the same questions, we examined the numerical differences between the responses to the two sets of questionnaires, using a paired sample t-test. We found that there were significant differences between the two sets of data regarding the exploration of unknown objects by the children using their hands ($p = 0.033$), and the verbal identification of at least one part of the child's body ($p = .033$). In addition, we also discovered that there were significant differences between the first and last session, for pointing to at least one part of their body when asked to do so ($p = .011$), and when identifying body parts in any way ($p = 0.041$). As comments, teachers added that one child has changed and that he is now able to listen and understand body parts. Another child changed to being more focused compared with his previous state and he was enjoying the body part activities much more.

B. Structured Interview

The interviewed teacher had prior knowledge about the robot's functionalities. During the interview the teacher classified the following behaviours as improved:

- "Child 1 held attention for a longer period of time";
- "When Child 3 touched KASPAR's face, he was completely engaged with the robot. And he was touching KASPAR's body parts and face, because he was happy. KASPAR was definitely facilitating the interaction between you [the experimenter] and Child 3, because he wanted to engage with you [experimenter]. His eye contact was just amazing.";
- "Child 5 was interacting, and I said previously I did not see him interacting with someone, but today he and Child 1 spoke to each other. And I stopped the lesson, for them to continue, because they were speaking to each other.".

C. Behavioural Analysis

To compare the data from the video analysis of the first and the last session we used a paired sample t-test. As mentioned above, one of the coded behaviours was the direction of the eye gaze of the children when they were interacting with the robot. We found significant differences, comparing the first and the last session, for the children looking at KASPAR ($p = .001$), at the experimenter ($p = .004$), and elsewhere ($p = .032$). The results (Fig. 5) show, that the average time the children looked at the robot decreased (75.04% - 51.01%), at the experimenter increased (4.29% - 16.01%) and to no particular place also increased (20.66% - 32.97%).

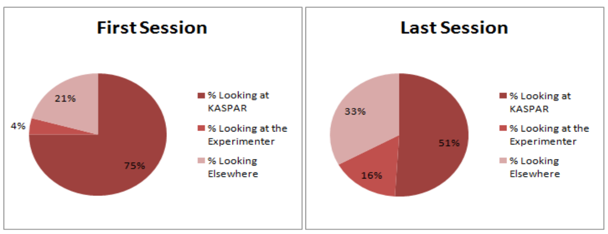


Figure 5. Percentage of eye gaze in the first and last session of the Practice Phase

Comparing the Group A (High Functioning) with Group B (Low Functioning), we found significant differences between the average time Group B looked at KASPAR in the first and in the last session ($p = .048$). There are no significant differences between the average of time that Group B looked at the experimenter or elsewhere. The average time the children in Group B looked at the experimenter increased from 4.32% to 15.3% group.

In Group A, we found significant differences in the average time of looking at KASPAR ($p = .025$), at the experimenter ($p = .033$), and looking elsewhere ($p = .417$), comparing the first and the last session. The average time that the children in Group A looked at the experimenter increased from 16.7% to 49.4%.

Concerning the tactile interaction of the children in the first and last session, there were no significant differences of the number of times the children touched the robot or the experimenter, gently ($p = .281$) or roughly ($p = .381$). Despite having no significant differences when evaluating tactile interaction, more than 90% of the times the children touched the robot gently (Fig. 6).

We video coded the following behaviours: the children following a pointing gesture with head movement of the experimenter (following), the children pointing at something with an index finger to attract the attention of the experimenter (pointing), and the children imitating vocalisations or gestures of KASPAR/experimenter (imitation). There are no significant differences between the first and the last session in any of the interaction parameters (pointing, following, and imitation). The behaviours that were shown most were imitation and pointing.

Regarding the success of the children while performing the

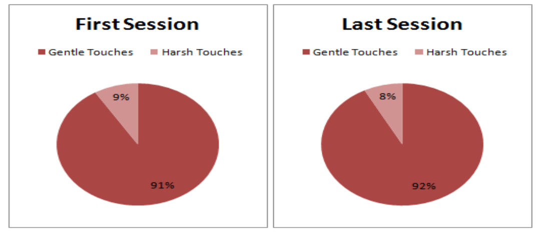


Figure 6. Percentage of gentle and rough touches during the interaction with KASPAR in the first and last session of the Practice Phase

proposed activities Figure 7 shows that the children managed to complete Activity A more than 70% of the times in both the first and the last session. There were no significant differences between the first and the last session.

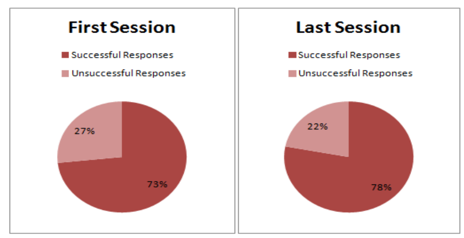


Figure 7. Percentage of Success of the Activity A in the first and last session of the Practice Phase

D. Comparison between pre- and post-test

When comparing the pre- and post-test there were no significant differences in the time children took to complete the performance task ($p = .365$). The average time the children took was 156 seconds in the pre-test and 124 seconds in post-test. 75% of the children managed to perform the task in less time in the post-test than in the pre-test.

V. Discussion

We observed that from the first session with KASPAR to the last, children directed their eye gaze increasingly less towards KASPAR. Instead the time they spent looking at the experimenter and at no particular place increased. The latter can be explained with the familiarisation of the children with the situation, but looking four times longer at the experimenter can be interpreted as KASPAR successfully functioning as social mediator. Comparing Group A and Group B, we found that the results follow the trend of the entire group. Despite the increase in time the children in Group B looked at the experimenter, they did so only 15% of the time in the last session. In Group A on the other hand this time increased up to almost 50% of the last session. The difference between the two groups show that children at the high functioning end of the autistic spectrum are much more attentive to the social partners face then children at the low functioning end. In both groups this attentiveness seemed to have been promoted by the activity with KASPAR. An increasing familiarity with

 ISBN: 978-1-61208-250-9

the experimenter could be an alternative explanation of these results, which nevertheless would be a desirable outcome of the triadic interaction between the robot, the experimenter and the child.

The fact that the difference between the first and the last session regarding the tactile interaction of the children with KASPAR was not significant, could have different reasons. One explanation could be that all the children were even in the first session gentle in more then 90% of their tactile interactions. This by itself is interesting, since the teachers reported that this initial gentleness was surprising to them. Based on this descriptive quantitative data it is possible to argue that the exposure of the children to the interactive situation with KASPAR already induced a more careful behaviour. However a more detailed evaluation of the tactile data from all sessions is needed to understand how the children interact with the robot.

Even without analysing quantitatively the number of times that the children performed interaction behaviours (pointing, following, and imitating), it is interesting to notice that imitation is the most pronounced behaviour. Previous studies with KASPAR [17] show that children on the high functioning end of the spectrum are able to imitate KASPAR's movements, and that it was easier for them to imitate and understand the partial movement of the body of KASPAR than the total movement of IROMEC, a mobile robotic platform. This indicates that KASPAR can be useful to facilitate interaction behaviours.

According to the data from the interview and the questionnaires some of the children that initially were not able to identify any of the body parts on themselves, showed an improvement on their knowledge. The teachers also indicated that the children transferred some of the knowledge learned during the sessions with KASPAR to the classroom. They gave in general very positive feedback

VI. Conclusion and Future Work

This paper presents a study in which children with autism interacted with a humanoid robot. The children learned about body parts and at the same time the robot was equipped to respond accordingly to tactile interaction from the children. We wanted to test whether the robot could facilitate the interaction between the child and another person in the experiment, and to acquire knowledge about human body parts. Another point of interest was to see if the robot could help children with ASD to learn appropriate physical social engagement. Our results show that from the first to the last session with the robot, the children increased the time they looked at the experimenter. An evaluation by teachers of the children shows, that they improved the ability to identify parts of the their body with their own hands. Additionally, some of the children that initially were not able to identify any of the body parts on themselves, showed an improvement of their knowledge. In this paper we used a robot to enable the learning of body parts by children with autism and due to the novelty of the topic, we wanted to construct and test different scenarios in this respect. A more detailed analysis of the data is still ongoing and will enable with a better interpretation of the findings. Due to the preliminary nature of the data analysis, no causal conclusions can be drawn at this point. It is necessary to point out that it is, due to the school environment design of the study, not possible to exclude that any observed improvements could be due to other activities at school or at home. Further research is needed to confirm the extent in which the robot was instrumental in causing these changes.

Acknowledgment

The authors are grateful to the professionals, and the students of the Southfield School, Hatfield, and their parents for their participation in the project. The authors are also grateful to the Portuguese Foundation (FCT) for funding through the R&D project RIPD/ADA/109407/2009 and SFRH/BD/71600/2010.

References

[1] A. Montagu, "Touching: The human significance of the skin," *New York*, 1971.
[2] M. Korkman, "Introduction to the special issue on normal neuropsychological development in the school-age years," *Developmental neuropsychology*, vol. 20, no. 1, pp. 325–330, 2001.
[3] J. Ratey and E. Hagerman, *Spark: The revolutionary new science of exercise and the brain*. Little, Brown and Company, 2008.
[4] N. A. for Sport and P. Education. (2002, feb) Naspe releases first ever physical activity guidelines for infants and toddlers. [Online]. Available: http://www.aahperd.org/naspe/template.cfm?template=toddlers.html
[5] R. L. Klatzky and S. J. Lederman, *Touch*, forth ed. New York: John Wiley & Sons, Inc., 2003, vol. 4, pp. 147–176.
[6] J. Bremner, *Infancy*. Wiley-Blackwell, 1994.
[7] M. K. Moore and A. N. Meltzoff, "New findings on object permanence: A developmental difference between two types of occlusion," *British Journal of Developmental Psychology*, vol. 17, no. 4, pp. 623–644, 1999. [Online]. Available: http://dx.doi.org/10.1348/026151099165410
[8] P. Smith, H. Cowie, and M. Blades, *Understanding children's development*. Wiley-Blackwell, 2003.
[9] K. Dautenhahn, C. Nehaniv, M. Walters, B. Robins, H. Kose-Bagci, N. Mirza, and M. Blow, "Kaspar–a minimally expressive humanoid robot for human–robot interaction research," *Applied Bionics and Biomechanics*, vol. 6, no. 3-4, pp. 369–397, 2009.
[10] B. Argall and A. Billard, "A survey of tactile human-robot interactions," *Robotics and Autonomous Systems*, vol. 58, no. 10, pp. 1159–1176, 2010.
[11] P. Marti, A. Pollini, A. Rullo, and T. Shibata, "Engaging with artificial pets," in *Proceedings of the 2005 annual conference on European association of cognitive ergonomics*. University of Athens, 2005, pp. 99–106.
[12] W. Stiehl, J. Lieberman, C. Breazeal, L. Basel, L. Lalla, and M. Wolf, "Design of a therapeutic robotic companion for relational, affective touch," in *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*. IEEE, 2005, pp. 408–415.
[13] A. Libin and E. Libin, "Person-robot interactions from the robopsychologists' point of view: the robotic psychology and robotherapy approach," *Proceedings of the IEEE*, vol. 92, no. 11, pp. 1789–1803, 2004.
[14] K. Wada and T. Shibata, "Robot therapy in a care house-its sociopsychological and physiological effects on the residents," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 3966–3971.
[15] B. Robins, F. Amirabdollahian, Z. Ji, and K. Dautenhahn, "Tactile interaction with a humanoid robot for children with autism: A case study analysis involving user requirements and results of an initial implementation," in *18th IEEE International Symposium on Robot and Human Interactive Communication RO-MAN*, 2010.
[16] G. Mesibov, V. Shea, and E. Schopler, *The TEACCH approach to autism spectrum disorders*. Springer, 2004.
[17] H. Lehmann, I. Iacono, B. Robins, P. Marti, and K. Dautenhahn, "'make it move': playing cause and effect games with a robot companion for children with cognitive disabilities," in *ECCE'11*, 2011, pp. 105–112.

An Interactive Game with a Robot: Peoples' Perceptions of Robot Faces and a Gesture-Based User Interface

Michael L. Walters
Adaptive Systems Research Group,
University of Hertfordshire,
Hatfield, Herts, UK,
AL10 9AB

M.L.Walters@herts.ac.uk

Samuel Marcos
CARTIF Foundation,
Division of Robotics and Computer Vision,
Parque Tecnológico de Boecillo,
205, 47151, Boecillo, Valladolid, Spain

sammar@cartif.es

Dag Sverre Syrdal & Kerstin Dautenhahn
Adaptive Systems Research Group,
University of Hertfordshire,
Hatfield, Herts, UK, AL10 9AB

D.S.Syrdal@herts.ac.uk
K.Dautenhahn@herts.ac.uk

***Abstract*—This paper presents findings from a HRI user-study which investigated participants' perceptions and experiences playing a simple version of the classic game, stone-paper-scissors with a humanoid robot. Participants experienced the robot displaying one of four different robot faces and interacted with the robots using a gesture-based interface. Findings from the study indicated that the effects of the different robot faces were inter-related with participants gender and ratings for overall enjoyment of the game experience. The usability and effectiveness of the gesture-based interface were overall rated positively by participants, though the use of a separate display for the game interface seems to have distracted participants attention from the robot's face.**

Keywords; HRI, Human-Robot-Interaction, Gesture-Based User Interface, Interactive Game Robot

I. INTRODUCTION

Recent research aims towards developing robots for use in domestic, office or other human-oriented environments. These robots will interact with humans in these environments as a matter of course and should exhibit behaviour that is not just safe and socially acceptable for humans in the vicinity, but should also facilitate and enhance the usability of the robots. Overall it has been found that people prefer consistency, in particular regarding robot appearance, capability and functionality [1][2]. The judgement of consistency of robot appearance, behaviour and function is subjective and it is likely that peoples' own idiosyncratic preferences and perceptions play a large role. However, some general characteristics regarding peoples preferences for robots have been discovered. Most people do not like robots to have realistic human appearances [3], but prefer some degree of human-likeness. A sizeable minority strongly prefer robots with non-human-like (i.e. machine-like) appearance [4][5]. It has been shown that people perceive human-likeness for robots as two main factors, physical-likeness and expressive-likeness, with the latter (i.e. communication and interaction abilities) generally more desirable than physical human-likeness (i.e. appearance) [6]. Most people would prefer to interact with a robot by means of speech [7][8] and many previous studies (e.g. [9][10][11]) have indicated that robot speech capabilities hold much promise with regards to Human Robot Interaction (HRI). However, speech recognition

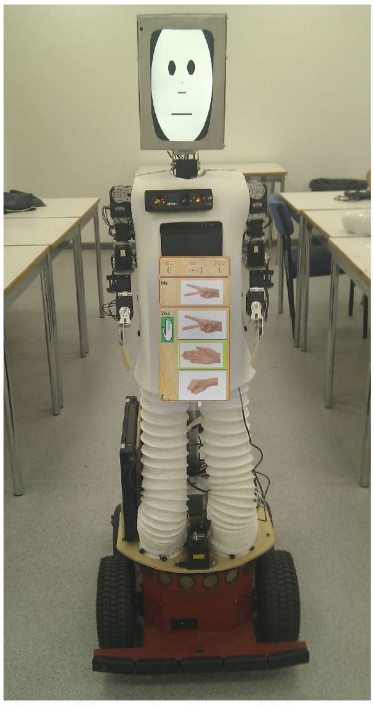


Figure 1. CHARLY (Companion Humanoid Autonomous Robot for Living with You) with 'retro robot' body covers and LCD face display showing a simple cartoon-like face.

technology is not yet reliable enough to be used by non-experts in human-oriented unstructured and noisy environments. Therefore, current robots are typically controlled by users using traditional keyboards, mouse, or touch based Graphic User Interfaces (GUIs). With the recent cost effective availability of the Microsoft Kinect [12], the use of gesture based user-interfaces has become a feasible way for users to interact with robots in a relatively natural way; either as a supplement to existing modes of interaction or as the main mode of interaction and control [13][14].

CHARLY (Companion Humanoid Autonomous Robot for Living with You) is a humanoid robot developed by the Adaptive Systems research group at the University of Hertfordshire (see Figure 1). It is particularly aimed at investigating robot appearance and social behaviour and HRI. During 2010-11 it was shown in public venues in project demonstrations, dissemination and as part of an art exhibit, "My Familiar Companion"

 ISBN: 978-1-61208-250-9

in collaboration with international artists, Anna Dumitriu and Alex May [15]. These venues were attended by relatively large numbers of people, which provided the opportunity to carry out survey-based research in conjunction with these events [6]. In the light of the findings, technical and procedural experience gained from these events, it was decided to carry out a new HRI study to investigate in more depth some of the issues that had arisen. The areas chosen for further investigation were: Peoples' perceptions and reactions to different face displays and to evaluate a newly developed Kinect gesture based user interface. This was to be achieved by playing a simple interactive game with the robot. The main research questions were:

1) Does the appearance of the robot's face display have any effects on users' perceptions or preferences for interacting with the robot?

2) Is the Kinect based gestural interface a feasible way for users to interact and control a robot.

Section II describes the robot and game system, Section III the user study method, Section IV presents the results and Section V the Discussion and Conclusions.

II. CHARLY SYSTEM DESCRIPTION

CHARLY has a simplified human-like appearance (humanoid) and scale, and was developed as part of the EU funded project LIREC (LIving with Robots and intEractive Companions). CHARLY is designed with easily replaceable head and body covers so that the robot's appearance may readily be changed to facilitate HRI research. CHARLY was developed using as many (low cost) standard parts as possible and the arms are primarily used for making expressive gestures. The body is mounted on two 'legs' and the robot can assume either a bending, crouching or sitting position. The robot body is mounted on a Pioneer 3AX mobile robot base [16].

1) CHARLY Server and Control

A dedicated Mini-PC implements a simple TC-IP based server which allows remote client programs to connect over a local (wired or wireless) LAN to control the robot. Remote clients can be written in any programming language that supports IP sockets. The underlying control software has obstacle detection and avoidance continuously active in normal use, so CHARLY has the capability to move safely in people orientated environments, either under direct control or autonomously.

2) Morphing Face Display System

A Face Morphing display system was developed when CHARLY was previously shown at the “My Familiar Companion” art exhibit [15]. This used a Microsoft Kinect camera mounted on the robot's chest to detect and isolate the faces of anyone coming within range. All the detected faces are combined into a single image such that faces closer to the Kinect have more influence over the resulting image than ones further away. The morphing face display program also acted as a client program for CHARLY and directly controlled the head and body movements for the “My Familiar Companion” exhibit.

3) CHARLY Stone Scissors Paper Game

For the current Study an Interactive Game program was developed which used the Kinect sensor to interpret users' gestures and implement a simple game based on Stone, Paper, Scissors. This program ran concurrently with the Face Morphing Program (only for one of the experimental conditions, see Section III) on CHARLY's laptop PC and allowed participants to interact with CHARLY using only gestures.

Initially it was planned to used CHARLY's hands to make appropriate gestures for Stone, paper or scissors, to use the Kinect to recognise hand gestures directly and to use speech to provide the main means of providing information to participants. However, though feasible in a quiet lab situation, it was found not robust enough in noisy and dynamic public areas. It was therefore decided to use a second body-mounted touch-pad to provide a dedicated display where the game progress and user interface was displayed (see Figure 3). Also, a method for controlling the game and robot with the human participants' whole arm gestures was found to be much more reliable in public environments. The program interfaced to the CHARLY Server and initiated simple pre-scripted body, head and arm movements, gestures and speech at appropriate points of the game.

4) Game Interface and Description

The game interface is shown in Figure 3. The topmost part displays the current game score. The robot score is labelled as “Me” and the user score is labelled as “You”. The middle area, also labelled as “Me”, displays an image of the robot's selection. During the game, this image changed randomly over time giving the user a clue as to which option the robot will choose. The bottom area has three buttons that allow the user to select rock, paper or scissors. This area is labelled as “You”. Each button has an image of the hand gesture for scissors, paper and rock as used in the real game. The user makes a choice by moving their (left or right) arm up and down (Figure 4) then pushing their hand forward when their desired choice icon is highlighted (Figure 5). When the user makes a selection, the image freezes to indicate the robot's choice. A complete game session consisted of three rounds, each of three individual games. The first player (robot or participant) to win two individual games in a round wins the round. The first to win two rounds wins the game. A complete game of three rounds typically lasted from 5 to 10 minutes. Depending on who won the game, the robot said “I won the game” or “You won the game” and either waved its arms in the air or adopted a head down posture respectively.

B. CHARLY Face Displays

The robot used four different face displays during the interactive game sessions. Each participant experienced only one face display from: *Simple static face* - A simple black and white, cartoon-like face constructed from ellipses and straight lines (Figure 6). *Morphing user face display* - The face display slowly changed to an image of the face of the the interacting person (see Figure 7 and section II.A.2 for more details). *A fixed image of a real robot face* - This robot face image was from a robot head we had used previously for live HRI studies [17][18] (Figure 8a). *Simple cartoon-like expressive face* - Three simple cartoon-like faces, constructed from ellipses and straight lines, formed simple 'expressions' which predicted the game intention of the robot (Figure 8b-c).

 ISBN: 978-1-61208-250-9

III. EXPERIMENTAL SET-UP AND PROCEDURE

The study ran for 2 weeks in late November 2011 and was performed in the foyer area of the University of Hertfordshire Learning Resources Centre (LRC). The LRC is open to students and staff working at the University and provides open access to research, library, computing resources and social areas. A high level of casual and passing foot traffic occurs within the foyer area and this provided a potentially large number of participants for taking part in the Interactive Game Study (IGS). CHARLY was set-up at one end of the entrance foyer area, with the robot operator, control and support equipment partially obscured by screens (see Figure 2). A separate desktop computer ran a browser-based questionnaire to obtain participants' views on their IGS experience. Brief descriptions of the questions are provided in the following section IV along with the questionnaire results and analyses. The experimental procedure was as follows:

1. Any passers-by that showed interest in the robot and game, were asked if they wanted to play a game with the robot. If the answer was positive, they were then asked for consent to take part in the study. Note, if they only wanted to play the game, this was acceptable. If they consented to take part in the study, they were asked to sign a consent form to allow data and/or video recordings to be collected and used for subsequent research purposes.
2. The game was described and the gesture-based interface was shown to them. After a short practice, the game was restarted and up three rounds were played. Many participants did not complete a full three rounds, either because a winner was declared after two rounds, or the participant wanted to finish.
3. The participant was then directed to the questionnaire computer and left to complete the online questionnaire. An experimenter was on hand in case of any difficulties with using the software, but did not lead or direct their responses to the questions.

IV. RESULTS

The questionnaire was divided into the following categories:

Demographic – Age, gender, occupation, handedness, robot computer and games experience, and Kinect (xbox) experience. The total number of participants was 80, 54 males (68%), 26 females (32%) and 7 (8%) were left handed. Their ages ranged from 17 to 60 (mean = 27), 40 (50%) were quite or very familiar with computer

Figure 2. A participant playing the interactive game with CHARLY in the University of Hertfordshire Learning Resources Centre

Figure 3. Stone-paper-scissors game interface when the user is being tracked and selecting paper

games but only 12 (15%) rated themselves as familiar with (toy or service) robots. 57 (71%) had not used a Kinect before.

Game Experience – Enjoy overall, like playing with CHARLY, who won?, like winning/losing, play again, understand game, game difficulty, like CHARLY as game partner and like game interface.

CHARLY specific – Overall appearance, looking at, pleasing, comfort, distraction, difficulty focusing on game. Face; like, annoy, prefer another? Voice; like, like talking?

A. Principal Component Analysis (PCA): Global Liking, Face Type and Demographics.

In order to categorise the response data, a series of PCA analyses were run. Variables loading less than .5 on any given factor were removed as part of the analyses. The predictors of Global Evaluation were assessed using a stepwise multiple regression. The variables entered into the initial model were: Gender, Age, Familiarity with Robots, Familiarity with the Xbox Kinect, Familiarity with Computer Games, Familiarity with Rock, Paper, Scissors Game, Face type (Simple Static Face, Simple Expressive Face, User Face and Robot Face). The final analyses are reported here:

1) Global Evaluation of the Experience

The PCA on the variables intended to investigate a general (global) evaluation of the interaction as a whole found a unidimensional construct, formed of three factors

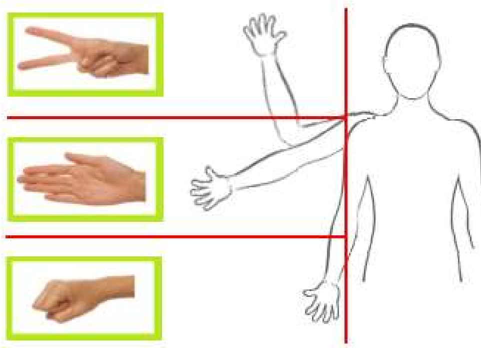


Figure 4. The three areas related to the user's body where a detected hand is associated with the selection of one of the three buttons of the game

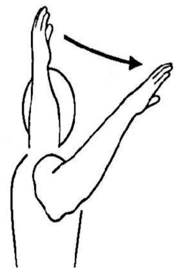


Figure 5. Schema of the movement necessary to press the scissors button

which had an eigenvalue of 2.22 and explained 73% of the variance. Table 1 provides the factor loadings and the variables are presented below:

Table 1: PCA of the Global Experience Evaluation Variables

Variable	Factor Loading
How was your experience of playing the game with CHARLY?	.87
Did you like playing with CHARLY?	.88
If you had the Opportunity, would you like to play with CHARLY again?	.83

2) Game Usability

Table 2: PCA of the Robot Evaluation variables

Variable	General	Voice	Distraction
Did you like CHARLY's General Appearance?	**.825**	.050	.083
Did you find it pleasing looking at CHARLY ?	**.839**	.284	.076
CHARLY distracted me from the game	.119	.029	**.828**
CHARLY made it difficult to focus	.010	.215	**.800**
Did you like CHARLY's face?	**.767**	.300	.116
I liked CHARLY's Voice	.015	**.853**	.057
Felt good talking with CHARLY	.123	**.843**	.105
Did you feel comfortable looking at CHARLY?	**.638**	.267	.360

The PCA on the variables intended to investigate the usability aspects of the game itself found a unidimensional construct formed of three variables which had an eigenvalue of 1.69 and explained 57% of the variance as follows:

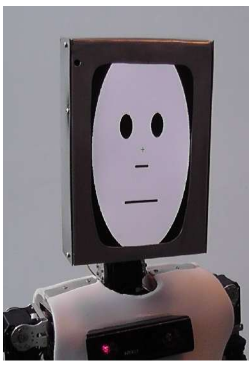


Figure 6. The Simple Static cartoon-like robot face display

Figure 7. The Morphing User face display

Table 3: PCA of the Robot Evaluation Variables

Variable	Factor Loading
Did you find it difficult to play the game?	.75
Did you like the interface of the game?	.65
Did you find it difficult to understand the game?	.84

3) Robot Evaluation.

The PCA on the variables intended to investigate the views of the robot in the interaction (Table 2) found three factors. The first factor had an eigenvalue of 2.63 and explained 33% of the variance. It was named General Robot Evaluation and is described below in terms of factor loadings. The Second Factor was named Robot Voice Evaluation with an eigenvalue of 1.77 and accounted for 22% of the variance. The Third Factor was named Robot Distraction with an eigenvalue of 1.24 and accounted for 16% of the variance. In total these three factors accounted for 70.5% of the variance.

B. Significant differences for the four face display conditions

There was no significant relationship between success at the game and the face display conditions ($\chi^2(3)=2.10$, $p=.553$), and also no relationship between success at the game and any of the Global, Game or Robot Measures.

ANOVA tests between the four face display condition groups' responses indicated that in terms of Global Enjoyment, there was a differentiation between the different facial displays. LSD post-hoc tests found that this difference was due to participants rating the Simple Expressive face as less enjoyable. However, there was also a significant interaction effect with face display type and gender ($F(3,71)=3.18$, $p=.029$). This relationship is illustrated in Figure 9 and suggests that male participants rated the Static Simple face display significantly higher

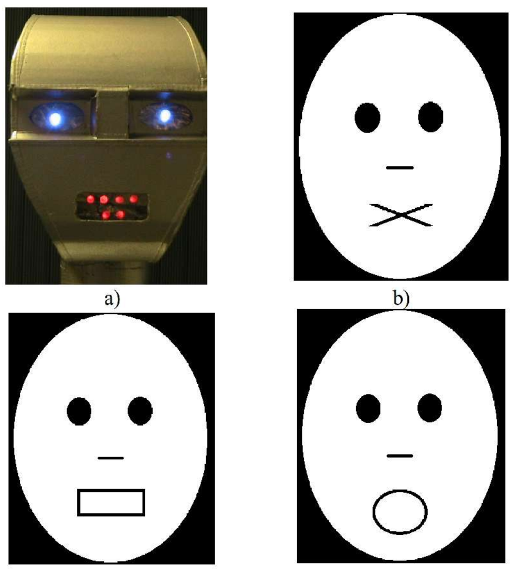


Figure 8. a) The real robot face display and Simple Expressive robot face which displays game intention; b) Scisssors, c) Paper and d) Stone.

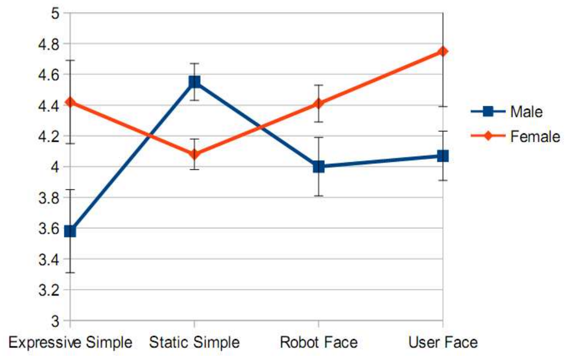


Figure 9. Interaction Effect for Gender and Face Display Conditions in terms of global evaluation.

than female participants and the opposite is true of the other display types. The effect was most pronounced for the Simple Expressive and the User Faces.

Table 4: ANOVA indicated significant differences between the four face conditions and participants' enjoyment of the interactive game experience (F(3,75=2.775, p=.047)

Group	N	Mean	Std Error
Simple Face	22	4.46	0.13
User Face	23	4.30	0.14
Robot Face	23	4.16	0.13
Expressive Face	12	3.86	0.22

There was no significant relationship between Face Display Condition along the other measures. There was, however a significant main effect for for Gender and the General Robot Evaluation measure (F(1,72)=4.27, p=.042). This effect suggested that female participants were more likely to rate the robot more positively along this dimension. The results also suggested that overall, participants rated the interaction positively.

C. Relationship between Prior Use of the Kinect Interface and Measurements.

There was no relationship between Prior Use of the Kinect and the Evaluation measures (F(1, 77)<.010, p>.99, η^2<.001) and no relationship between Prior Use of the Kinect and Success at the Game. In fact, there was a non-significant trend in which participants with no experience of the Kinect were *more* likely to succeed (χ^2(1)=.522, p=.470).

Table 5: Correlation Matrix for Evaluation Measures

Measure	Global	Game	Robot General	Robot Voice	Robot Distract
Global	1				
Game	**.197***	1			
Robot General	**.254****	**.380****	1		
Robot Voice	.191	-.039	.088	1	
Robot Distract	.104	**-.287****	.171	-.141	
Note: *p<.01, **p<.05					

This combined with the overall high scores given for Evaluation, suggest that the gestural interface was suitable for a wide section of participants, not just those accustomed to such interfaces from previous use.

D. Correlations between the Measures

The different measures were correlated using Pearson's r, and the results are presented in Table 5 which suggests that there are results approaching significance between Global Evaluation, Game Evaluation and the General Robot Evaluation measures. In addition, the Distraction measure was negatively correlated with the Game Evaluation.

V. DISCUSSION AND CONCLUSION

Regarding the first research question: While there was a difference between the Face Display conditions in terms of Global Evaluation, post-hoc tests found that this was caused by the sample as a whole rating the simple expressive display less favourably than the other displays.

Significant gender differences in the sample were found, with female participants being more positively inclined towards the Expressive, Robot and User Face conditions than the male participants. This suggests that male participants overall preferred a simple, unchanging face display, while female participants were more positive towards a richer, more dynamic display. This finding echoes those from [19] where male participants had a tendency to prefer robot behaviours that facilitated the explicit goal of the interaction, while female participants attached more value to the more social and intrinsically rewarding aspects of the interaction. This also mirrors results found in [20] where female players of games seemed to value the non-competitive aspects of computer-mediated games in comparison with male players. Also, the female participants may have been able to process the richer facial displays more efficiently than male participants as found in [21]. This suggests that while some participants found it problematic to deal with split attention in such interactions, the use of rich and dynamic face displays will still be evaluated. However, the simple expressive face which predicted the robot's next game move was not highly rated by participants, and it is therefore likely that they were so concentrated on the game display, that they did not pay much attention to the face display. The problem of split attention has been remarked upon in some previous studies [22][23][24] and these findings seem to reflect a similar phenomenon.

For the second research question the data supports using the Kinect based gestural interface as a feasible way to interact and control the robot. Although half the participant sample had previous computer game experience, most had no previous experience of playing a Kinect-based (computer) game. Even though participants' interaction time was typically less than 10 minutes, their ratings for usability were relatively high irrespective of their previous experience of the Kinect interface.

This study has shown that there are differences in how participants evaluate the use of a robot. It has highlighted the trade-off in the current system between the potential problem of split attention with competing sources of expressive and game task feedback. Differences were found between male and female perceptions' of the interaction; game interaction and control, versus the intrinsic reward of a richer, more dynamic face display. Future work should address the issues of incorporating a gestural interface integrated with social feedback mechanisms in a more seamless manner for service robots in human-oriented environments.

ACKNOWLEDGMENT

Many thanks are due to Anna Dumitriu and Alex May for their work in developing the morphing face system and body form for CHARLY. This work was supported by the LIREC (LIving with Robots and intEractional Companions) EU project (FP7-215554), the Junta de Castilla y León (CCTT/10/VA/001/A.2-C.6), the EDRF programme and the Spanish Ministry of Economy and Competitiveness (DPI2011-25489, DPI2008-06739-C02-01, IPT2011-0854-900000).

REFERENCES

[1] A. Tapus, M. J. Matarić, "User Personality Matching with Hands-Off Robot for Post-Stroke Rehabilitation Therapy" in *Proceedings of the 10th International Symposium on Experimental Robotics (ISER-06)*. 2006.

[2] S. Lee, I. Yee-man Lau, S. Kiesler, C. Chiu, "Human Models of Humanoid Robots" in *Proceedings of the 2005 International Conference on Robotics and Automation (ICRA 05)*. 2005. pp. 2767- 2772

[3] T. Minato, K. F. MacDorman, M. Shimada, S. Itakura, K. Lee, H. Ishiguro, "Evaluating Humanlikeness by Comparing Responses Elicited by an Android and a Person" in *Proceedings of the Second International Workshop on Man-Machine Symbiotic Systems.*. 2004. pp. 373-383

[4] M. L. Walters, D. S. Syrdal, K. L. Koay, K. Dautenhahn, R. te Boekhorst, "Human Preferences and Perceptions of Robot Appearances" in *Proceedings of Artificial Intelligence and Simulation of Behaviour Convention (AISB 09), HRI Symposium*. 2009. pp. 136-143

[5] D. S. Syrdal, M. L. Walters, K. L. Koay, S. N. Woods, K. Dautenhahn, "Looking Good? Appearance Preferences and Robot Personality Inferences at Zero Acquaintance" in *AAAI - Spring Symposium 2007, Multidisciplinary Collaboration for Socially Assistive Robotics*. 2007. pp. 86-92

[6] M. L. Walters, D. S. Syrdal, K. Dautenhahn, A. Dumitriu, A. May, B. Christiansen, K. L. Koay, "My Familiar Robot Companion: Preferences and Perceptions of CHARLY, a Companion Humanoid Autonomous Robot for Living with You" in *In Proceeedings of the Conference of Towards Autonomous Robotics Systems (TAROS 2012)*. 2012.

[7] K. Dautenhahn, S. N. Woods, C. Kaouri, M. L. Walters, K. L. Koay, I. Werry, "What is a Robot companion - Friend, Assistant or Butler?" in *Proceedings of IEEE RSJ International Conference on Intelligent Robot Systems (IROS'05)*. 2005. pp. 1488-1493

[8] B. Bengtsson, J. K. Burgoon, C. Cederberg, J. Bonito, M. Lundeberg, "The Impact of Anthropomorphic Interfaces on Influence,Understanding, and Credibility" in *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences-Volume 1*. 1999. pp. 1051-1066

[9] S. Lauria, G. Bugmann, T. Kyriacou, E. Klein, "Mobile Robot Programming Using Natural Language." *Robotics and Autonomous Systems*, 2002, 38 (3-4). pp. 171-181

[10] C. Nass, S. Brave, *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. The MIT Press, 2005.

[11] A. J. N. van Breeman, K. Crucq, B. J. A. Krose, "A User-Interface Robot for Ambient Intelligent Environments" in *Proceedings of ASER 2003*. 2003. pp. 132-139

[12] Microsoft Kinect. http://www.xbox.com/en-US/kinect.

[13] H. B. Suay, S. Chernova, "Humanoid Robot Control Using Depth Camera" in *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2011)* . 2011. pp. 401-402

[14] S. M. Albrektsen, Thesis. Using the Kinect Sensor for Social Robotics. Norwegian University of Science and Technology, Faculty of Information Technology, Mathematics and Electrical Engineering, Department of Engineering Cybernetics. 2011. http://urn.kb.se/resolve?urn=urn:nbn:no:ntnu:diva-13862

[15] My Robot Companion. An art exhibit commisioned by the Science Gallery, Dublin, and created in collaboration with the University of Hertfordshire. http://www.myrobotcompanion.com/.

[16] Robotis Inc. Manufacturers of the Dynamixel series of robot servos. www.robotis.com.

[17] Mobile Robots. http://www.mobilerobots.com/.

[18] The OpenNI API Framework. www.openni.org.

[19] Open Computer Vision Library (OpenCV). http://sourceforge.net/projects/opencvlibrary.

[20] M. L. Walters, K. Dautenhahn, R. te Boekhorst, K. L. Koay, "Exploring the Design Space of Robot Appearance and Behaviour in an Attention-Seeking 'Living Room' Scenario for a Robot Companion" in *Proceedings of IEEE-Artificial Life (Alife 07)*. 2007. pp. 341-347

[21] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst, K. L. Koay, "Avoiding the Uncanny Valley – Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion." *Journal of Autonomous Robots*, 2008, 24(2). pp. 159-178

[22] H. Kose-Bagci, K. Dautenhahn, D. S. Syrdal, C. L. Nehaniv, "Drum-mate: Interaction Dynamics and Gestures in Human-Humanoid Drumming experiments." *Connection Science*, 2010, 22(2). pp. 103-134

[23] T. Hartmann, C. Klimmt, "Gender and Computer Games: Exploring Females' Dislikes." *Journal of Computer-Mediated Communication*, 2006, 11(4). pp. Article 2

[24] E. Hampson, S. M. Anders, L. I. Mullin, "A Female Advantage in the Recognition of Emotional Facial Expressions: Test of an Evolutionary Hypothesis." *Evolution and Human Behavior*, 2006, 27. pp. 401-416

[25] M. L. Walters, M. Lohse, M. Hanheide, B. Wrede, D. S. Syrdal, K. L. Koay, A. Green, H. Hüttenrauch, K. Dautenhahn, G. Sagerer, K. Severinson-Eklundh, "Evaluating the Robot Personality and Verbal Behavior of Domestic Robots Using Video-Based Studies." *Advanced Robotics*, 2011, 25(18). pp. 2233-2254

[26] M. Lohse, M. Hanheide, B. Wrede, M. L. Walters, K. L. Koay, D. S. Syrdal, A. Green, H. Hüttenrauch, K. Dautenhahn, G. Sagerer, K. Severinson-Eklundh, "Evaluating extrovert and introvert behaviour of a domestic robot – a video study" in *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (RoMan 08)*. 2008. pp. 488-493

[27] G. Zen, B. Lepri, E. Ricci, O. Lanz, "Space speaks: towards socially and personality aware visual surveillance" in *Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis*. 2010. pp. 37-42

Robust Perception of an Interaction Partner Using Depth Information

Salah Saleh, Anne Kickton, Jochen Hirth and Karsten Berns
Robotics Research Lab, Dept. of Computer Science
University of Kaiserslautern
Kaiserslautern, Germany
Email: {saleh,a_kickto,hirth,berns}@cs.uni-kl.de

***Abstract*—Social interactive robots require sophisticated perception abilities to behave and interact in a natural way. The proper perception of their human interaction partners plays a crucial role. The reduction of the false positive rate for human detection is very important for increasing the natural interaction abilities. This paper presents a combined method using RGB data as well as depth information to find humans in the robot's surrounding. To track a person over time a Kalman filter is applied, which also reduces the processing time. Furthermore, a head pose estimation on the basis of Support Vector Machines is integrated, which can be used to perceive nonverbal expressive cues like nodding. The proposed method is tested in various experiments.**

***Keywords*-social robots; perception; human-robot interaction;**

Figure 1: The humanoid robot ROMAN of the University of Kaiserslautern playing a question-answer game.

I. INTRODUCTION

Inspired by an ever older growing population in recent years the development of service and assistance systems for applications like elderly care, nursery, or entertainment and edutainment gained an enormous importance in the field of robotics. Therefore, they need special interaction capabilities that enable socially acceptable behavior. The University of Kaiserslautern is developing a humanoid robot ROMAN for investigating social human-robot interaction, Figure 1. Its control architecture is realized based on psychological principles to enable intrinsic motivation to fulfill specific tasks as well as emotional reactions depending on the progress of the current interaction [1]. To test and evaluate the behavior of the robot, it has been enabled to handle various interaction scenarios.

Interactive robots need the ability to realize complex combinations of speech, gestures, facial expressions, and body poses. One of the main aspects for the realization of human-robot interaction scenarios is a stable perception of the human interaction partners. The challenging task of the perception system is to analyze and understand sensors data. One of the most important sensors to fulfill this task is the vision system, which provides a large data stream. The interactive robot has to analyze this stream of data to understand the surrounding environment as well as the interaction partner. Starting with face detection, the present work proposes head pose detection and a tracking using RGB and depth information. There are many challenges in finding faces in a stream of images due to the variation in scale and orientation, pose, facial expressions, and light conditions [2]. Most of the research works focused on using two-dimensional images. Because of the limitations of the 2D features to describe the reality, the faces detected using only RGB images do not always represent real faces.

In recent years, most researchers focused on face detection using depth information. *Niese et al.* [3] have applied a model-based matching for the study of facial features and the description of their dynamic changes in image sequences. Their face detection is based on color driven clustering of 3D points derived from stereo image sequences. The face detection and normalization method consists of six steps. In the first step they calibrate the stereo camera. In the second step they generate a surface model of the individual face from an active stereo scan and save it in the database. In the third step, they process sequentially the stereo image sequence with a passive stereo algorithm. The fourth step localizes and post processes the face with help of 3D and color information. In the fifth step, they determine the face pose on the basis of an Iterative Closest Point algorithm (ICP) that finds the matching between the calculated surface model from the database and the post-processed data from stereo data. The last step creates a synthetic image of the face in frontal pose and standardized size. *Burgin et al.* [4] believe that face in video stream can be made more efficient by tracking the detected face. This fact minimizes the search space in each frame by searching only a local neighborhood around faces found in previous frames. They also utilize the depth of each pixel to calculate the possible size of the face centered on the pixel. They restrict their search for the faces of specific size in each depth. *Lu Xia et al.* [5] proposed

 ISBN: 978-1-61208-250-9

a method for human detection using depth information by Kinect. Their method is a model based approach. It detects humans using a 2D head contour model and a 3D head surface model. They proposed a segmentation scheme to discriminate the human from the background. They used a two-stage head detection process. The first stage explores the boundary information to locate the candidate region that may contain a human. In this stage they used 2D chamfer distance matching. This matching scheme scans the whole image to give the possible regions that may contain a human. These regions are then examined using a 3D head model using the relational depth information of the array for verification. The parameters of the head are then extracted from the depth array and used to build a 3D head model. The second stage matches the 3D model against all the detected regions to make a final estimation. They have also developed a region growing algorithm to find the entire body of the human.

An overview of different approaches for head pose estimation can be found in [6]. These approaches can be categorized depending on the type of data they used. Most research focused on using 2D images, while some of them focused on using depth information.

Seemann et al. [7] have proposed a neural network-based system using depth information. They have used a variation of color-based face detection techniques in addition to the depth information for fast and reliable face detection. They have used a three layers feed-forward neural network trained by back-propagation algorithm on a depth information obtained by the stereo camera.

Breitenstein et al. [8] have presented a real-time algorithm to estimate the 3D pose of a previously unseen face from a single range image. They have generated an average 3D face model from the mean of an eigenvalue decomposition of laser scans of 138 adults. They start with finding the nose tip and its orientation to roughly estimate the head pose. They used a 3D shape signature which is computed for each pixel to identify noses. The 3D positions and mean orientations of the nose candidate pixels form a set of head pose hypotheses. They used an error function in evaluation the alignment of reference pose range images and the input range image to estimate the pose. This process is performed in parallel on the GPU.

Fanelli et al. [9] have proposed a real time head estimation with random regression forests using depth data. Their training examples consist of range images of faces annotated with 3D nose location and head rotation angles. They assumed that the head has been already detected in an image. They constructed each tree of the forest from a set of fixed size patches randomly sampled from the training examples. Each patch consists of the extracted visual features associated with the pose parameters of that patch.

This paper presents an approach using depth information to overcome the problem of inaccuracy in human perception. The depth information is used in addition to RGB images to verify the detected faces – e.g. it can be distinguished from a photo. This improves the detection process by decreasing the false positive rate. The Kalman filter is used for face tracking as well as for decreasing the process time of the face detection. It predicts the position of the face in the next frame to reduce the search space. The goal of the face detection and tracking process is to follow the human face and record its movements to determine the persons emotion state.

The remainder of this paper is organized as follows: In Section II some of related concepts will be presented. Section III provides more detailed information on the implementation of the proposed system. Some experiments will be presented in Section IV. In Section V a conclusion and future work will be discussed.

II. CONCEPT

In addition to its cameras, a humanoid robot usually has multiple sensors especially those that provide depth information. These additional sensors can be used to make the detection of human beings, and by doing so also the interaction, more efficient and more accurate. This paper presents an approach using depth information in addition to the RGB images to perceive the interaction partner by detecting his/her face position, orientation, and movements.

A. Face Detection

One of the essential skills of the robot interacting with humans is face detection. In computer vision terms, the face detection task is not easy even though that humans can do it effortlessly [10]. The goal of face detection is to determine whether or not there are any faces in a given image, and if present, returns the location and size of each face [2]. It is a very challenging task for the robots, and has been one of the most studied research topics in the past few decades. The difficulty associated with face detection can be attributed to many variations in scale, location, orientation, pose, facial expression, lighting conditions, occlusions, etc. A faster and more reliable face detection process is a basic condition for proper human-robot interaction.

The face detection field has made significant progress in the last 20 years. One of the most important works in this field is the novel work of Viola and Jones [11]. Most of the face detection approaches have focused on the use of two-dimensional images without any additional information [2][10]. The additional sensors found in most robots such as depth information sensor can provide additional information that the robot can use in efficient and accurate face detection [4]. Depth information has several advantages over 2D intensity images. It is simple representations of 3D information in addition to its robustness to the change in color and illumination [5].

The present paper uses the RGB image and depth information provided by Kinect sensor. The detection process is accomplished in two stages. The first stage uses Haar cascade classifier to detect faces (candidate faces) in 2D image. The second stage takes the candidate faces detected by the first stage and uses depth information to verify if they are really faces or not.

B. Head Pose Estimation

In addition to speech, people have the ability of interacting nonverbally using different aspects. One of these nonverbal aspects is the human head movement. Humans have the ability of interpreting these movements quickly and effortlessly, while it is regarded as a difficult challenge in computer systems and robotics. Detecting the human head movement requires estimating the head pose (position and orientation) over the time. In a computer vision context, the process of detecting the pose of a human head from digital imagery is called head pose estimation. In order to build a robust human-robot interaction system, a robust and reliable head pose estimation algorithm is needed.

In addition to the sensitivity to illumination, the head pose estimation in 2D images suffer from the lack of features due to occlusion (in some poses) [8]. The depth information is used in the last years and it displayed encouraging results. The present paper uses the depth information in head pose estimation. Three Support Vector Machines (SVMs) for regression are trained to detect the pose angles (pitch, yaw and roll). After detecting the face position, the pose estimator applies the depth features of the face to the SVMs to estimate the pose.

C. Face Tracking

The face tracking is another important aspect of human-robot interaction in order for designing a robust interaction. In this paper, the Kalman filter [12] is used for face tracking. The kalman filter also used to speed up the face detection process and to give more confidence to the pose estimation. The prediction of the face position, size and orientation in the next frame enable us to search within a specific area of the frame. This reduces the detection time tremendously.

The Kalman filter is an algorithm which uses a series of measurements observed over time and produces estimates of unknown variable. The motion state of the face can be formulated by the following state model:

$$x_k = Fx_{k-1} + Bu_k + w_k \quad (1)$$

Where

- x_k is the state vector in time k.
- F is the state transition which is applied to the previous state x_{k-1}.
- B is the control input model which is applied to control vector u_k.
- w_k is the process noise.

At the time k, a measurement z_k of the state x_k is calculated according to the following model:

$$z_k = H_k x_k + v_k \quad (2)$$

Where

- z_k is the observation (measurement) vector.
- H_k is the observation model which maps the true state space into the observed space.
- v_k is the observation noise.

Based on the above two models, the state vector along with its covariance matrix can be updated to predict the next position and orientation of the face.

III. IMPLEMENTATION

The perception of human using 2D images encounters serious problems because its sensitivity to illumination and shadow. The recent 3D acquisition systems could help overcoming these problems. Schmitz and Berns [13] have suggested a communication partner model. They have used auditory and visual perception system. The present paper proposes a robust visual perception system for human-robot interaction using depth information. The system uses depth information in addition to the standard 2D images in perceiving the partner. The system uses Kalman filter to track the face position and orientation to keep eye contact as well as reducing the calculation time. An overview of the system modules is depicted in Figure 2.

In training phase, a large database of 15K, 640x480 range images of faces for 20 persons has been used [9].

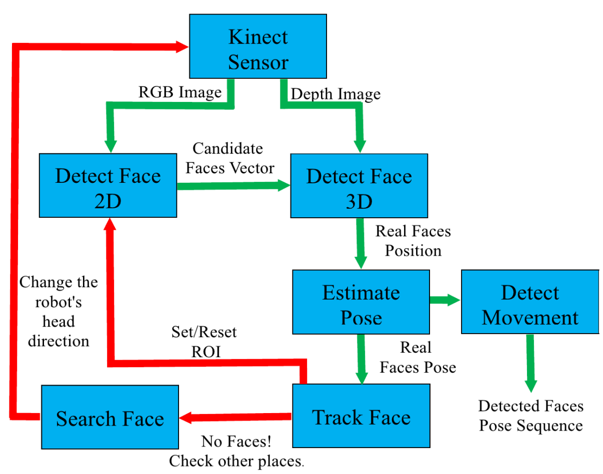


Figure 2: The proposed system block diagram.

A. Face Detection

The face detected by Haar cascade classifier does not always represent a real face. This is because that this method uses 2D features that can be found in any variations of intensities (colors) that may similar to a human face. Also it detects each face in a picture as a real face what represent a problem at the level of human-robot interaction. This problem can be overcome using depth information. The faces detected using cascade classifier can be checked again to determine if they are really faces or not. This can improve the detection process by decreasing the false positive rate.

In this paper, the 2D detection module uses multi-appearance cascade classifier to detect the face. Then the 3D detection module verifies the detected faces and decides which of these candidate faces are real faces and which are not. In this context, the detected face using the cascade classifier referred to as *candidate face*. The input of this module is a list of candidate faces with their positions. Its output is a list of detected faces with less false positive rate. The verification process examines each of the candidate faces. The corresponding depth information of each candidate face is used rather than the RGB information.

The candidate face is simply rejected if it doesn't meet one of the following criteria:

1) The relation between the depth (the distance from the face to the camera) and the face size must be governed by the following equation as presented by [5].

$$h = p_1.z^3 + p_2.z^2 + p_3.z + p_4 \tag{3}$$

Where

- z is the depth of the center of the face,
- h is the size of the face,
- $p_1 = -1.3835 \times 10^{-9}$
- $p_2 = 1.8435 \times 10^{-5}$
- $p_3 = -0.091403$
- $p_4 = 189.38$

If the candidate face doesn't satisfy the above formula then the face will be rejected and considered as not real.

2) The candidate face must have a reasonable depth to be a real face. The low depth or high depth face may represent a picture of a face or a color variation in the image. The face depth is calculated as the difference between the maximum and minimum distances after removing the outliers.

$$facedepth = max\_depth - min\_depth \tag{4}$$

If the depth of a candidate face is less than a predefined threshold or greater than a predefined threshold then the face is rejected and considered as not real.

3) The face must not be included in another face. If there is such case, then the face nearest to the criterion 1 is regarded and the other is omitted.

B. Head Pose Estimation

Head pose estimation is an essential skill that is needed in human behavior analysis. Estimating a head pose using depth information has shown very reasonable results. The proposed work uses depth information in head pose estimation.

Three SVMs for regression [14] are used as head pose estimators for the three angles *pitch*, *yaw* and *roll*. These SVMs are trained to find the pose angles from a set of depth features. These features are calculated after normalizing the depth information of the face. By dividing the face into set of rows and columns and omitting the borders, these features are the average depths of the regions lie on these rows and columns.

After calculating the depth features of the face, the Principal Component Analysis (PCA) is performed to find the most important features. This process reduces the problem dimensionality and speeds up the training of SVMs. Then the SVMs are trained using the resulting features from the PCA.

When a new face is applied, the depth features of the face are calculated and the most important features are derived using the eigenvectors to be supplied to the SVMs.

C. Face Tracking

Human face tracking is important aspect in order of natural interaction. A Kalman filter is used for human face tracking. It also decreases the process time of face detection. It predicts the position and orientation of the face in the next frame to reduce the search space. It uses twelve variables as state vector components. The six variables (x, y, z, pitch, yaw and roll) represent the face position in 3D space and three orientation angles. The other six variables are the velocities (vx, vy, vz, vpitch, vyaw and vroll). The (x,y) point is used to determine the position of the face in the next frame whereas the z value (depth) is used to determine the estimated face size according to equation 3. Depending on the face position and size in the next frame, a Region Of Interest (ROI) is determined to reduce the search space. The ROI will be searched to detect the face in the next frame rather than the whole frame. This will reduce detection time tremendously. The position of the ROI is centered on the position of the face in the next frame. While the size of the ROI is calculated by multiplying the face size by some *factor* to ensure that the whole face is located within the ROI.

To find a suitable face size factor for the tracking algorithm, empirical studies have been conducted. Figure 3 shows the relation between the number of frames per second and the face size factor for each depth. Figure 4 shows the optimal face size factor for each depth. The linear equation we get from the experiments is:

$$factor = 0.00053 \times depth + 0.8667 \tag{5}$$

Algorithm 1 shows the process of detection, pose estimating and tracking faces.

IV. Experiments

To evaluate the proposed system different experiments have been conducted. These experiments proved reliability within distance range 1-4 meters.

 ISBN: 978-1-61208-250-9

Algorithm 1 Face Detection, Head Pose Estimation and Tracking

Initialize Kalman Filter;
Set the Region Of Interest (ROI) as the whole frame ;
for each frame from the sensor (RGB and Depth) **do**
  Detect Face in the ROI using Cascade Classifier (RGB);
  Get a list of candidate faces positions (x,y) and sizes;
  Verify the detected faces using depth information (z);
  Get a list of the real faces positions and the depth (x,y,z);
  if there is real faces **then**
    Report the existence of a human face at the position (x,y,z);
    Estimate the head pose;
    Record the head pose information;
    Use the current position (x,y,z) and the current pose (pitch,yaw,roll) as the state variables in Kalman filter;
    Predict the next position of the face (x′,y′,z′) and the next pose (pitch′,yaw′,roll′);
    Get the face size in the next frame from the value of z′ using equation 3;
    Set the position of the ROI position (x′,y′) in the next frame;
    Calculate the face size factor from equation 5;
    Set the size of the ROI as $(facesize * facesizefactor)$;
  else
    Move the robot's head to search for a face depending on the last position of the face and the motion direction;
    Set the ROI as the whole frame;
  end if
end for

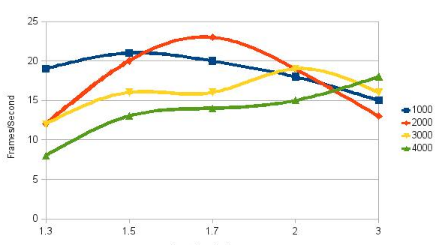


Figure 3: The relation between the frames per second and the face size factor.

Experiment 1

In order to assess the proposed face detector, a real-time stream of data from Kinect sensor with resolution of 640x480 has been examined. The experiment has shown that the false positive rate of the proposed method is lower than using only 2D data as depicted in Figure 5. The cascade classifier scans the entire image searching about faces with many scales as a first stage. The second stage does not need to rescan the entire image. It only examines the detected faces from the previous stage. Consequently, the time needed to examine faces in the second stage is not noticeable. The detected faces were in a rate of 5 frames per second. The experiment has also shown that the Haar cascade classifier influences the perception system especially in the bad lighting conditions.

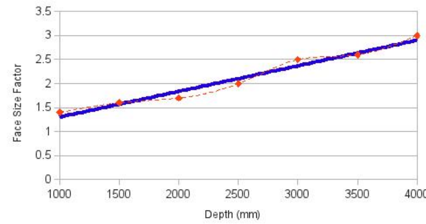


Figure 4: The optimal face size factor for each depth.

(a) candidate faces = 6 (b) real faces = 2

Figure 5: 2D vs. 3D face detection

Experiment 2

This experiment is to compare the proposed head pose estimation with other work. ETH Face Pose Range Image Data Set [8] has been used for comparison. It contains over 10K range images of 20 peoples in many different poses. The head pose ranges cover about $\pm 90°$ yaw and $\pm 45°$ pitch rotations. The parameters of the three SVMs are different. Some of them are set by conducting some experiments. The experiments have shown precise and fast responses. Table I compares the results of the proposed head pose estimator with two other methods. It shows the mean and standard deviation of the pose estimation errors in addition to the percentage of correctly classified poses based on $10°$ threshold.

Table I: Comparison of the proposed head pose estimation method with [8] and [9]

	Yaw error mean/stdev	Pitch error mean/stdev	Pose Estimation Accuracy
The proposed method	$3.7/3.4°$	$2.6/3.0°$	94.9%
Fanelli et al.	$5.7/15.2°$	$5.1/4.9°$	90.4%
Breitenstein et al.	$6.1/10.3°$	$4.2/3.9°$	80.8%

Experiment 3

This experiment demonstrate the use of Kalman filter in human tracking. It reduces the processing time tremendously. Figure 6 shows the processing time with and without tracking. It has shown a processing speed average of 17 frames per second compared to 5 frames per second without tracking. Using the depth in the state variables of Kalman filter overcomes the problem of tracking two persons when they cross each other. Also tracking the head pose increases the estimation confidence.

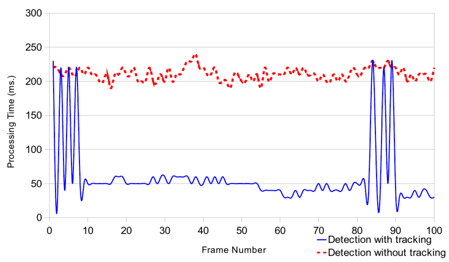


Figure 6: Processing time

V. Conclusion and Future Work

Social interactive robot require stable method for detecting interaction partners. For natural interactive behavior it is important to minimize the false positive rate of these algorithms. The paper presented an approach using a combination of RGB images and depth information to improve the human detection. Besides the pure perception the presented approach also provides a possibility to determine the pose of the head using the 3D information. For tracking a person's face over time a Kalman filter has been added, which also reduces the processing time of the proposed algorithm. This also enables to detect nonverbal cues like shaking the head or nodding. The developed approach has been evaluated in an interactive game scenario.

The next steps in this context will be to include further experiments to evaluate the quality of the developed system and its application for human-robot interaction. Furthermore, the system should be extended in a way that facial expressions as well as body postures of the interaction partners can be evaluated to gather information on the interaction partners emotional state. Other nonverbal expressions like conciously performed gestures should be also recognized in order to improve the robot's interactive capabilities.

References

[1] J. Hirth, "Towards socially interactive robots – designing an emotion-based control architecture," Ph.D. dissertation, Department of Computer Science, University of Kaiserslautern, April 16 2012.

[2] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, Jan 2002, pp. 34–58.

[3] R. Niese, A. Al-Hamadi, and B. Michaelis, "A novel method for 3d face detection and normalization," *Journal of Multimedia*, vol. 2, no. 5, 2007.

[4] W. Burgin, C. Pantofaruy, and W. Smart, "Using depth information to improve face detection," in *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*, march 2011, pp. 119 –120.

[5] L. Xia, C.-C. Chen, and J. Aggarwal, "Human detection using depth information by kinect," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, june 2011, pp. 15 –22.

[6] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation in computer vision: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 4, pp. 607 –626, April 2009.

[7] E. Seemann, K. Nickel, and R. Stiefelhagen, "Head pose estimation using stereo vision for human-robot interaction," in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, may 2004, pp. 626 – 631.

[8] M. Breitenstein, D. Kuettel, T. Weise, L. van Gool, and H. Pfister, "Real-time face pose estimation from single range images," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, june 2008, pp. 1 –8.

[9] G. Fanelli, J. Gall, and L. V. Gool, "Real time head pose estimation with random regression forests," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, june 2011, pp. 617 –624.

[10] E. Hjelmas and B. K. Low, "Face detection: A survey," in *Computer Vision and Image Understanding*, vol. 83, no. 3, Sept. 2001, pp. 236–274.

[11] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[12] G. Welch and G. Bishop, *An Introduction to the Kalman Filter*, University of North Carolina Chapel Hill, Chapel Hill (NC), USA, July 2006.

[13] N. Schmitz and K. Berns, "Perception systems for naturally interacting humanoid robots," in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Atlanta, GA, USA, July 31 - August 3 2011, pp. 425–432.

[14] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, aug 2004.

Gesture Recognition for Humanoid Assisted Interactive Sign Language Tutoring

Bekir Sıtkı ERTUĞRUL, Cemal GURPINAR, Hasan KIVRAK, Ajla KULAGLIC, Hatice KOSE
Department of Computer Engineering
Istanbul Technical University
Istanbul, TURKEY
{bsertugrul, gurpinarcemal, hkivrak , kulaglic, hatice.kose }@itu.edu.tr

***Abstract*— This work is part of an ongoing work for sign language tutoring with imitation based turn-taking and interaction games (iSign) with humanoid robots and children with communication impairments. The paper focuses on the extension of the game, mainly for children with autism. Autism Spectrum Disorder (ASD) involves communication impairments, limited social interaction, and limited imagination. Many such children show interest in robots and find them engaging. Robots can facilitate social interaction between the child and teacher. In this work, a Nao H25 Humanoid robot assisted the human teacher to teach some signs and basic upper torso actions which were observed and imitated by the participants. Kinect camera based system was used to recognize the signs and other actions, and the robot gave visual and audial feedback to the participants based on the performance.**

Keywords-Human-robot interaction; autism; imitation games; sign language

I. INTRODUCTION

Communication is a vital requirement for human life. Language acquisition is an extremely crucial process for brain development and intelligence. Sign Language (SL) is an alternative way of communication for hearing impaired or autistic children who cannot communicate verbally. Sign Language is a visual language that is based on upper body movements (including hands, fingers, arms, upper torso, head and neck) and facial gestures. There are studies on visual recognition of sign language, and sign language tutoring with 2-D visual aids to hearing impaired people [1-6]. Also several robots and robotic hands were utilized to implement sign language [7-8]. In the studies [9-11] visual games are employed for sign language tutoring.

The studies introduced in this paper have been realized as part of an on-going research, which aims to utilize humanoid robots for assisting sign language tutoring due the lack of sufficient educational material. Also in terms of children's sign language education 2-D instructional tools are found to be incompetent. Therefore using humanoid robots as an assistive tool in sign language tutoring for children will be very beneficial. In the proposed system, it is intended that a child-sized humanoid robot is going to perform and recognize various elementary signs (currently basic upper torso gestures and words from sign languages) so as to assist teaching these signs to children with communication problems. This will be achieved through interaction games based on non-verbal communication, turn-taking and imitation that are designed specifically for robot and child to play together. We have used imitation based non-verbal interaction games with humanoid robots successfully with children and adults previously in [12-16].

Currently, American Sign Language (ASL) and Turkish Sign Language (TSL) are being implemented and tested within the project.

In the first versions of the game, the robot was telling a short story verbally and through the story for some selected words, the robot was able to express word in the SL among a set of chosen words using hand movements, body and face gestures and having comprehended the word, the child was encouraged to give relevant feedback in SL or visually to the robot (using a colored card visualizing the word), according to the context of the game. The games were demonstrated with more than 100 preschool children with hearing ability and 7 preschool children from Special School for Hearing Impaired Children [17-22] (game demo videos: http://humanoid.ce.itu.edu.tr/).

The current paper summarizes the attempt to extend this study to autistic children as a part of a PhD course entitled as Autism and Computational Aspects, which aims to bring researchers in computer science and robotics with non-engineer experts from psychology, neurology, speech therapy, and autism therapists, to train students who want to work in this field, and for a long term brain storming event on possible computational solutions that can be used within autism therapy (recognition of autism and other related issues were not included in the course schedule due to time limitations). This paper presents one of the projects which are produced as an output of this collaboration, and it is planned to use the system and the game in the collaborative special schools on autism. Autism Spectrum Disorder (ASD) involves communication impairments, limited social interaction, and limited imagination. Researchers are interested in using robots in treating children with ASD [23-27]. Many such children show interest in robots and find them engaging. Robots can facilitate interaction between the child and teacher. Every child with autism has different needs. Robot behavior needs to be changed to accommodate individual children's needs and as each individual child makes progress.

The game will be based on the visual cards, the cards will be shown to the robot to select among several signs from ASL and basic upper torso motion (hands side, forward, up etc.) Then the robot will perform the sign and wait for the child to imitate. The imitated action will be evaluated using an RGB-D camera (Kinect) and robot will give a motivating comment when the action is imitated with success.

 ISBN: 978-1-61208-250-9

Figure 1 - Some of the cards used in these exercises

II. Sign Recognition

A. Hidden Markov Model

Every hidden state in Hidden Markov Model (HMM) which models the hand motion is responsible for a specific part of given symbol sequence. In homogeneous hidden Markov models, the durations of segments are modeled with geometric distribution. These durations for every state are independent from each other. This constraint becomes important while types of hand movement and the number of different user increase.

The generative models, like HMM, model the common probabilities between observations and states. In distinctive models, the probability distributions of states depend on the observations and it does not model the probability distributions of the observations subject to classes. It is hoped that distinctive models are more efficient in classification. Thus, in order to recognize hand movement, Conditional Random Fields (CRF), which is equivalent to HMM and its kinds were examined [28].

B. Hidden Conditional Random Fields(HCRF)

CRF's do not model the internal dynamics of the class, but inter-class dynamics and because of this constraint it is not suitable for the classification of time series. Therefore Hidden Conditional Random Field (HCRF) [29] and Hidden Dynamic Conditional Random Field (HDCRF) [30] were offered. While HDCRF models both the internal and external dynamics of classes, HCRF only models the internal dynamics and therefore it is more convenient for the isolated hand movement recognition problem.

HCRF's associate the observations with the state transitions instead of learning the state durations. This attribute increases the performance of positive samples, but it also increases the false acceptance rate.

C. Input Output Hidden Markov Models(IOHMM)

Input-Output HMMs as generative and discrete hybrid models show high performance in recognizing hand movements [31]. In IOHMM like HCRF, state transition probability distribution depends on the input sequence that consists of the function of observations [32].

In IOHMM, observations and state transition probabilities are calculated from input sequence using local models. Radial basis functions or multi-layer perceptron's can be used for local models. IOHMM's are more complicated than HMMs and their training requires more samples than HMMs.

D. Hidden Semi-Markov Models(HSMM)

Although HSMMs are similar to the HMMs. HSMMs hidden states produces observation sequence from certain probability distributions instead of producing a single observation [33,34]. HSMM state creates a symbol sequence instead of a single symbol. Fixed-Term Models (FTM) is a kind of HSMM and it determines the exact staying duration at each state with the status of a plug-in counter. FTM solves the problem of the modeling of specific periods, but durations are still independent of each other unless being conditioned to the velocity and dimension.

In order to recognize isolated hand movement, every action class must be modeled from positive samples. When an unclassified hand movement sequence comes, it is evaluated by all defined models in system and class likelihoods are calculated and the class of a model with the highest value is selected as a label. At [35], for evaluating the performance ratio and recognition rate of HMM, IOHMM, HCRF and FTM data set is collected. To ensure the independence of recognition rates from vision modules, a Kinect camera with infrared sensor is used. As a result of the experiments, it was found that performance rates of HCRF and IOHMM are higher than FTMs and HMMs, but they are slow for real time systems. For real time systems, FTMs that shows high performance more than HSMMs are offered.

III. Proposed Method

In the data collection phase, RGBD camera (Kinect Sensor) starts the input stream and sends every gestural motion data in the form of frame by frame. Different gestural motion data taken from RGBD camera (Kinect Sensor) can have different number of frames. Thus representation of every gesture pattern should be carefully modeled so that recognition process meets the performance criteria for robust recognition.

As a first step, joint spatial coordinates (x,y,z) of skeletal structure for each joint are generated. Then, in order to provide robustness, every frame in the gestural motion data is expressed as a single vector of angle values (Roll, Pitch, Yaw) which is computed from spatial position values (x,y,z) for every joint node of skeletal model. Several image processing and computer vision techniques are used to detect the skeletal model of human successfully using RGBD cameras i.e. Kinect Sensor. Thus the job on image processing in determining a good feature for gesture classification became easier with the availability of Kinect.

There are two main goals for this study. First, one is to represent gesture pattern (Sign Language (SL) word) using a suitable classification algorithm (K-Means). It provides clusters from every spatial motion data coming from the Kinect sensor corresponding to the related centroid of data. The generated probabilistic model (Hidden Markov Model) which is generated as a result of the system accepts this clustered discrete data.

In the second phase, the recognizer cycle is started to provide recognition of this gesture so that recognized gesture patterns (SL words) are adaptively transferred to humanoid robot (Nao). In order to recognize the gesture, it generates a dynamic model for every distinct behavior (gesture). According to the clustered data coming from the K-Means algorithm, it determines hidden states (node) and observable variables (output labels). In the training section, data as a target vector (a collection of observation sequences) seeds into recognizer cycle to perform supervised training algorithm (e.g. Baum-Welch). Finally, recognizer model throws a unique distinct behavior as a label (related SL word/ gesture).

IV. INTERACTIVE ISIGN GAME

As stated before, as an outcome of the PhD level course on Autism and Computational Aspects, an imitation game was constructed using Kinect and Nao H25, based on the recommendations of autism therapists. Our main goal was to extent our studies on signing based interaction games with humanoid robots, to autism therapy.

Figure 2 - “up”, and “side” actions.

The sign imitation game is an extension of the sign language game; as an initial step, we used basic upper torso gestures, i.e. opening the arms sides, up, forward, waving hand, etc., in the long run, we plan to use signs from ASL and TSL as well. The aim of the game is to teach the children to recognize and imitate the gestures/signs, within a turn-taking interaction game. The demonstrator will be the robot and the therapist will be able to manually assist the child, when the child fails to imitate the action successfully. Within this game it is possible to locate many of the exercises already being used as a part of the autism therapy.

The game consists of 3 stages. In the first stage the child will learn how to play doing the gestures one by one, the sequence and the quality of the gestures were chosen by the therapist or the child. When they show a picture of the gesture to the robot, the robot does the gesture and waits for the child to repeat the action. Using a Kinect camera we can evaluate child’s actions and send the robot feedback. If the child can repeat the action then robot says “you did the action good”(The experts suggested us that we have to praise the action of the child, it is not enough to say “its good” or “congradulations”). Else the therapist helps the child manually to do the gesture. (The experts told us we should not let the child do the action wrong, because then the action will be learned wrong).

In the second stage, the game is like a sports work out, each action/gesture is repeated several times without the picture display and the therapist get involved less. The child is assumed to learn each action by now.

In the third stage, we turned the game into a musical play. Robot sings a song related to the actions and do the actions one by one and the child is expected to repeat the sequence of actions.

Figure 3 - Demo setup with participant, Nao robot and Kinect Device (a second Kinect device was also employed for finger detection but not used within the demos)

Figure 4 - Therapists help children in the first stage of the game

The robot will record the success rate of the child and also the experimenter will record the therapist’s corrections and child’s success.

These games were usually played with the therapist, or the video of the therapist, or another autistic child. The robot will act as a play mate in these games.

V. GAME IMPLEMENTATION

A. Data Collection

Kinect and Microsoft Kinect SDK v1.5 was used to get human joint data from human motion. The human skeletal model of Kinect camera system is used to get 20 different joint data (X, Y, Z coordinates) over the SDK. Only upper torso information (shoulder, elbow, wrist coordinates) is used in this study, to classify the actions. Neck, head, face and fingers were no included within this study, but will be included in the overall project.

As a start, 6 actions, namely, 3 basic sign language action (ASL and TSL) and 3 basic upper torso actions from 15 university students were recorded with Kinect. For each action, every participant was asked to perform 4 times, of which 2 were used for training the system and 2 were used for testing. 30 frames/sec were recorded by the camera. Discrete HMM was used for sign recognition. The actions to be implemented were selected specially based on the advises of the therapists so that the tests can be also applied to the children with autism or mental disabilities. Also the actions which can be implemented by the Nao robot (due to Kinematic constraints) were selected among these advised actions. In order to avoid boredom, confusion and fatique in children during the tests, number of the actions and the repeats were set to a minimum, as possible.

The actions were introduced to half of the participants (6 participants) using the real physical robot and the simulated robot, and by a real human to the rest of the participants. All of the participants were asked to stay on a special sign on the floor and implement the signs to make sure that they keep the same distance to the Kinect camera during the tests (1.5 mt). The participants were not given any information except the fact that they should follow the robot/human and imitate the actions afterwards, and their actions are recorded with the Kinect camera.

First of all, a classification algorithm was used to enable the usage of discrete observation symbols and states on the motion data captured by Kinect. K-Means method was used to discretize the motion data and classified, then for every motion, in total 6 HMM was trained and the parameters were determined with Baum-Welch algorithm.

Afterwards, the observation sequence was matched to the related action using the 6 different HMM which were trained before, and their probability to produce these observations was calculated using forward algorithm. The action matching the biggest probability was classified with this action. If it was trained with the same model in the training set, then the classification was successful. The complexity of the HMM should be detected according to the data used with the model.

B. Aldebaran(NAO) Tools Used

In this project, the goal is to empower teacher that works with children with ASD and to customize robot behavior to suit the needs of each child. NAO H-25 humanoid robot is used during the field studies, since it is a small size humanoid robot which is suitable to implement basic signs in the ASL and TSL, robust and safe to work with children. For further studies a bigger size humanoid robot platform with 5 fingers and more DOF on arms will be used within the project.

Aldebaran Robotics offer several software tools for use with the NAO robot. Choregraphe can be used for face detection, face recognition, speech, speech recognition, walking, recognizing special marks and dances, and individual control of the robot's joints. The movements can be performed in sequence or in parallel. Choregraphe needs to be used with a robot proxy, real or simulated.

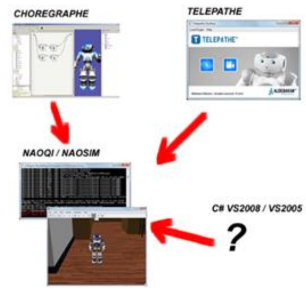


Figure 5 - Interaction of Aldebaran software for the NAO robot. Owners can use various software tools including Visual Studio to develop their own NAO software.

The simulated proxy can be NAOqi or a sophisticated simulator such as NaoSim [36]. NAOqi is a piece of software that simulates the robot for Choreographe and tests it before trying on the actual robot. NAOsim is simulation tool that allows for robot simulation in an apartment having simulated furniture. Whereas NAOqi only simulates the robot, NAOsim simulates an environment with which the robot can interact. Monitor, which was called Telepathe until the current release, allows the user to access the robots memory, see through the robots two cameras and observe the environment as the robot senses it. Also, it is possible to use some of program languages, Python or C++ to program the NAO.

During the implementation phase, different methods for the kinematic modelling of the humanoid robots are available such as [38].

C. Image Recognition

To navigate through the exercises, the instructor used a set of cards with different images. Each image represents a different exercise. This requires image recognition software which was coded using C++ and OpenCV 2.3.1. The algorithms used in this project included SURF feature detection, Bag-of-Words, K-means clustering and support vector machines (SVM).

OpenCV has many algorithms to detect and describe local feature. SURF was chosen for feature extraction and object detection. This is one of the most common methods.

The Bag of words (BoW) approach is more commonly used in natural language processing. When used for image processing, an image and its graphical features are analogous to the document and to the words, respectively. After image representations had been obtained with BoW, SVM supervised learning was used for classification. SVM takes an input array which consists of data and a label. The label represents the class to which data belongs. There are many hyperplanes that could be used to classify the data sets. The best hyperplane is the one that gives the greatest separation or margin between the classes.

The training database consisted of a scanned set of Walt Disney cartoon characters printed on cards. Every card was identified by a number which became the SVM label for that card. Training began by extracting the features from the images and computing image descriptors:

Then, the dictionary of graphical features was determined with K-means clustering as shown below. The result set of BowKmeansTrainer was written to file in YAML format. That improves speed of image recognition. The last step in training was training the SVM. CvSVM, an implementation included in OpenCV was used. The SVM training results were also written in a file to improve performance.

Once training was complete, the cartoon characters printed on the cards can be recognised by simply detecting the features of the images and sending them to the SVM for prediction.

D. Test and Result

6 Hidden Markov Models belonging to 6 signs were trained with 219 training sample and tested with 80 test samples. In this test every class has different state and event counts. The data which belongs to these classes, clustered corresponding to these event and state counts. The confusion matrix showing the tests with these parameters is shown in the Table 1. These state and event counts are the ones which gave the best solution in the test trials. At the tests, 8 state and 10 event were used for "side", 8 state, 10 event were used for "forward", 7 state, 9 event were used for "table", 4 state, 5 event were used for "car", 4 state, 6 event were used for "up", 6 state, 9 event were used for "dad".

The successfully matched classes were located in the diagonal of the table. 54 of the 80 test samples were classified successfully. When the unsuccessful groups were studied, it is observed that this is caused by the groups with similar features (similar actions).

For example "table" was recognised as "forward" 1 times, and "dad" and "up" signs were mislabeled. Another reason for the mislabeled signs, is that, some participants realized some actions wrong (not similar to the human/humanoid teacher), partly or as a whole.

TABLE I. CLASSIFICATION MATRIX

		Predicted classification					
		Araba (Car)	Baba (Dad)	Forward	Masa (Table)	Side	Up
Actual classification	Araba (Car)	13	0	0	1	1	0
	Baba(Dad)	0	11	1	0	1	1
	Forward	0	0	8	0	0	5
	Masa(Table)	3	1	1	7	0	0
	Side	3	1	3	0	6	0
	Up	1	0	2	1	0	9

In the future works, it is planned to improve the system recognition performance using probabilistic machine learning methods such as Hidden Conditional Random Fields, Input-Output Hidden Markov Models and Hidden Semi Markov Models.

There were no statistically significant difference between the performance of the actions imitated from human teacher and humanoid teacher (simulated/physical).

Within the studies, 1 child with normal developments, and 2 children with ASD (all in the age group 6-7) participated, as well as the university students. The children did not stay stable during the recording of data, which makes it very hard to get data. The child with the normal development and one of the children with ASD could finish all the actions, yet not exactly same with the adults. Unfortunately, since they did not stay still and were tired and leave the test before enough data was collected, the Kinect system could not be trained to recognize their actions. We are working on the necessary improvements to get data from children as fast and robust as possible. The system should be fast, and the time for calibration should be as short as possible.

VI. CONCLUSION AND FUTURE WORK

In this paper, we introduced a project on teaching children sign language by means of interaction games with humanoid robots. The extended versions of the studies are also being used with children with autism in teaching non-verbal communication skills, imitation and turn-taking. Several types of media including robots with different embodiment, tablets, and web based applications are being used within the study. The experiments are being conducted with adults, sign language students, children with normal development, hearing impaired children and children with autism. The main aim of this interdisciplinary study is to build a bridge between the technical know-how and robotic hardware with the know-how from different disciplines to produce useful solutions for children with communication problems. Moreover we would like to increase the awareness among families and public.

ACKNOWLEDGMENT

This work was supported by The Scientific and Technological Research Council of Turkey under the contract TUBITAK KARIYER 111E283. Author Bekir S. Ertugrul is supported by TUBITAK BILGEM.

REFERENCES

[1] Staner, A. T., and A. Pentland, "*Real-Time American Sign Language Recognition from Video using Hidden Markov Models*", Technical Report TR-306, Media Lab, MIT.

[2] Kadous, W., "*GRASP: Recognition of Australian Sign Language Using Instrumented Gloves*," MSc. Thesis, University of New South Wales, 1995.

[3] Murakami, K. and H., Taguchi, "*Gesture Recognition Using Recurrent Neural Networks*," Proceedings of CHI'91 Human Factors in Computing Systems, 1991, pp. 237-242.

[4] Aran, O., and L. Akarun, ―A Multi-class Classification Strategy for Fisher Scores: Application to Signer Independent Sign Language Recognition‖, Pattern Recognition, Vol. 43, no. 5, pp. 1717-1992, May 2010.

[5] Aran, O., I., Ari, P. Campr, E. Dikici, M. Hruz, S. Parlak, L.Akarun, and M. Saraclar, Speech and Sliding Text Aided Sign Retrieval from Hearing Impaired Sign News Videos , Journal on Multimodal User Interfaces, vol. 2, n. 1, Springer, 2008.

[6] Caplier, A, .S. Stillittano, O. Aran, L. Akarun, G. Bailly, D. Beautemps, N. Aboutabit, and T. Burger, Image and video for hearing impaired people, EURASIP Journal on Image and Video Processing, Special Issue on Image and Video Processing for Disability, 2007.

[7] Jaffe, D., Evolution of mechanical fingerspelling hands, for people who are deaf-blind, Journal of Rehabilitation Research and Development, 31, 236–244. 1994

[8] Hersh, M. A. and M. A. Johnson (Eds.), Assistive Technology for the Hearing-impaired, Deaf and Deaf blind, 2003, XIX, 319 p. 168 illus., Hardcover, ISBN: 978-1-85233-382-9

[9] Adamo-Villani, N. A virtual learning environment for deaf children: Design and evaluation. IJASET – International Journal of Applied Science, Engineering, and Technology, 16, 18-23, 2006

[10] Lee, S., V. Henderson, H. Hamilton, T. Starner, H. Brashear, and S. Hamilton A Gesture-based American Sign Language (ASL) Tutor for Deaf Children, Proceedings of CHI(Computer-Human Interaction). Portland, OR. April 2005

[11] Greenbacker, C., and K. McCoy. The ICICLE Project: An Overiew. First Annual Computer Science Research Day, Department of Computer & Information Sciences, University of Delaware, Feb 2008.

[12] Shen, Q., J. Saunders, H. Kose-Bagci, K. Dautenhahn, "An Experimental Investigation of Interference Effects in Human-Humanoid Interaction Games", Proceedings of IEEE RO-MAN2009 , pp. 291 - 298

[13] Kose-Bagci, H., K. Dautenhahn, C. L. Nehaniv, "Emergent Dynamics of Turn-Taking Interaction in Drumming Games with a Humanoid Robot", Proceedings of. IEEE RO-MAN 2008, pp. 346 – 353.

[14] Kose-Bagci, H., E. Ferrari, K. Dautenhahn, D. S. Syrdal, and C. L. Nehaniv, "Effects of Embodiment and Gestures on Social Interaction in Drumming Games with a Humanoid Robot", Special issue on Robot and Human Interactive Communication, Advanced Robotics vol.24, no.14, 2009.

[15] Kose-Bagci, H, K. Dautenhahn, D. S. Syrdal, and C. L. Nehaniv, "Drum-mate: interaction dynamics and gestures in human–humanoid drumming experiments," Connection Science, vol. 22, no. 2, pp. 103– 134, 2010.

[16] Dautenhahn, K., C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. A. Mirza, M. Blow KASPAR - A Minimally Expressive Humanoid Robot for Human-Robot Interaction Research. Special Issue on "Humanoid Robots", Applied Bionics and Biomechanics 6(3): 369-397, 2009

[17] Kose, H., R. Yorganci , H. E. Algan, and D.S. Syrdal, "Evaluation of the Robot Assisted Sign Language Tutoring using video-based studies", SORO special issue on "Measuring Human-Robot Interaction, 2012, DOI: 10.1007/s12369-012-0142-2

[18] Kose, H., R. Yorganci , and H. E. Algan, "Evaluation of the Robot Sign Language Tutor using video-based studies", Proceedings of 5th European Conference on Mobile Robots (ECMR11), *pp. 109-114*, 7-9 September, 2011.

[19] Kose, H., and R. Yorganci , "Tale of a robot: Humanoid Robot Assisted Sign Language Tutoring", 11th IEEE-RAS International Conference on Humanoid Robots, Bled, Slovenia (HUMANOIDS 2011) , pp 105 – 111, 2011,

[20] Kose, H., R. Yorganci , and I.I. Itauma, "Robot Assisted Interactive Sign Language Tutoring Game", IEEE ROBIO Conference, Thailand, pp. 2247-2249, 7-11 december, 2011

[21] Ertugrul, B.S., H. Kivrak, E. Daglarli, A. Kulaglic, A. Tekelioglu, S. Kavak, A. Ozkul, R. Yorgancı, H. Kose, "iSign: Interaction Games for Humanoid Assisted Sign Language Tutoring", International Workshop on Human-Agent Interaction (iHAI 2012), 11 October, 2012, Vilamoura, Algarve, Portugal, *Accepted.*

[22] Kivrak,H., B.S. Ertugrul, R. Yorgancı E. Daglarli, A. Kulaglic, , H. Kose, , Humanoid Assisted Sign Language Tutoring, 5th International Workshop on Human-Friendly Robotics, Brussels, October, 2012 *Accepted*

[23] Feil-Seifer, D. J., M. P. Black, E. Flores, A. B. St. Clair,E. K. Mower, C. Lee, M. J. Mataric, S. Narayanan, C. Lajonchere, P. Mundy, and M. Williams. Development of socially assistive robots for children with autism spectrum disorders. Technical Report CRES-09-001, USC Interaction Lab, Los Angeles, CA, 2009.

[24] Goodrich, M. A., M. A. Colton, B. Brinton, and M. Fujiki.A case for low-dose robotics in autism therapy. In Proceedings of the 6th international conference on Human-robot interaction, HRI'11, pp: 143–144, New York, NY, USA, 2011.

[25] Billard A., Robins B., Nadel J., Dautenhahn K. Building Robota, a Mini-Humaoid Robot for the Rehabilitation of Children with Autism. RESNA Assistive Technology Journal, vol.19, 2006

[26] Dautenhahn K. (2003) Roles and Functions of Robots in Human Society - Implications from Research in Autism Therapy. Robotica 21(4), pp. 443-452.

[27] Kozima H., C. Nakagawa, and Y. Yasuda. Children-robot interaction: a pilot study in autism therapy. Prog Brain Res, 164:385–400, 2007.

[28] Lafferty J. D., A. McCallum, ve F. C. N. Pereira (2001), "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in Proceedings of the Eighteenth International Conference on Machine Learning. 2001, pp. 282–289,

[29] Wang S. B., A. Quattoni, L. Morency, ve D. Demirdjian (2006), "Hidden conditional random fields for gesture recognition," Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, USA, 2006, pp. 1521–1527

[30] Morency L., A. Quattoni, ve T. Darrell (2007), "Latent-dynamic discriminative models for continuous gesture recognition," Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, vol. 0, pp. 1–8, 2007.

[31] Keskin C. ve L. Akarun (2009), "Stars: Sign tracking and recognition system using input-output HMMs," Pattern Recogn. Lett., vol. 30, pp. 1086–1095, September 2009.

[32] Bengio Y. ve P. Frasconi (1996), "Input-output HMM's for sequence processing," IEEE Transactions on Neural Networks, vol. 7, no. 5, pp. 1231–1249, September 1996.

[33] Yu S. ve H. Kobayashi (2006), "Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden markov model," IEEE Transactions on Signal Processing, vol. 54, no. 5, pp. 1947–1951, 2006.

[34] Murphy K. P. (2002), "Hidden semi-markov models," M.I.T. Technical Report, 2002.

[35] Keskin C. (2011), Belirli süre modelleri ile izole el hareketi tanıma", IEEE Sinyal İşleme ve İletişim Uygulamaları Konferansı, Antalya, 2011.

[36] Aldebaran Robotics Choregraphe. [Online]. http://www.aldebaran-robotics.com/en/ , Available: 13.02.2012

[37] Wang J.G., and Y. Li , "A Cooperated-Robot Arm Used for Rehabilitation Treatment with Hybrid Impedance Control Method", Intelligent Robotics and Applications, Eds by H. Liu, H. Ding, Z. Xiong and X. Zhu, LNCS 6425, 2010, pp.451-462.

Knowledge-driven User Activity Recognition for a Smart House. Development and Validation of a Generic and Low-Cost, Resource-Efficient System

Ismael Duque, Kerstin Dautenhahn, Kheng Lee Koay, Ian Willcock and Bruce Christianson
Adaptive Systems Research Group, School of Computer Science
University of Hertfordshire
Hatfield, Hertfordshire, UK
Email: {i.duque-garcia2, k.dautenhahn, k.l.koay, i.willcock, b.christianson}@herts.ac.uk

***Abstract*—Our core interest is the development of autonomous and socially interactive robots that may support elderly users at home as part of a smart home, i.e. a home equipped with a sensor network that may detect activities of daily living such as preparing food in the kitchen, having meal in the living room, watching the television, etc. The current paper focuses on showing the design and implementation of a low-cost, resource-efficient activity recognition system that can detect user activities without the necessity of collecting a large dataset to train the system. Based on common-sense knowledge from activities of daily living, we generated a set of rules for defining user's activities in a home setting. These rules can be edited and adapted easily in order to accommodate different environments and daily life routines. The approach has been validated empirically with a pilot study in the University of Hertfordshire Robot House. The paper presents results from a study with 14 participants performing different daily life activities in the house. The results are promising, and future work will include the integration of this system in a Smart House used for Human-Robot Interaction studies. This may help develop context-aware robot companions capable of making better decisions to support users in their daily activities.**

***Keywords*-Activity Recognition; Smart Houses; Context-Aware**

I. INTRODUCTION

In the field of Human-Robot Interaction (HRI), many researchers are interested in understanding how humans interact with robots in different environments [1]. The incorporation of social skills into robots' responses to achieve smoother interaction with humans remains a significant challenge. Many studies (e.g. [2] [3] [4]) from the Adaptive Systems Research Group at University of Hertfordshire have been carried out with the aim of gathering findings that help us understand how people interact with robots in a domestic environment, and hence to develop robots which exhibit a greater awareness of context when interacting with humans. The Robot House (see Figure 1) is the naturalistic environment used by our research group to perform this variety of experiments.

Fong et al. [5] assume that humans tend to interact with robots in ways that are similar to how they interact with other humans, i.e. humans expect certain social characteristics from robots. For instance, in the area of assistive robotics, the robots will become part of people's lives, so these social skills have to be enhanced during interaction in order to increase robots' acceptance in these environments. Context-aware robot companions would have the ability to detect users' activities performed at home, but they require additional modules such as human activity recognition systems. These supply the necessary information to allow robots to adapt their behaviour to the ongoing activity, and increase their social skills aforementioned.

One of the current problems pointed out in the literature regarding these systems, is the large variety of datasets necessary to create accurate activity recognition systems [6], and the difficulties in recruiting participants for the experiments [7]. We therefore developed a different method to avoid involving users in extensive studies of data collection during the whole process of system development. This point is particularly important when working with elderly people or people with special needs, which are one of the target user groups that our research is concerned with. Asking e.g. elderly people to spend several days or weeks engaged in certain activities to generate training data for the system puts a huge burden on them. The Activity Recognition System (ARS) that will be presented in this paper takes into account this issue. The knowledge-driven approach [8] used allowed us to develop the low-cost, resource-efficient system, in which participants were involved just during the validation stage.

Our research follows two well-defined directions. Firstly, the incorporation of social skills in robot companions to create more natural human-robot interactions in living environments. Smart homes' facilities will help to develop these skills (e.g. the non-intrusive sensor network installed in the Robot House). As Chan et al. [9] mentioned, sensor-embedded houses provide context information without disturbing users' daily activities, creating greater comfort and well-being. Secondly, we avoid the involvement of users during the training phase of the development of the system by the use of knowledge-driven approach. Following these two directions, we have created a functional activity recognition system that was tested with 14 participants in its validation stage.

The remainder of this paper is organised as follows: Section 2 discusses related work. Section 3 presents the research question and goals. Section 4 describes how the activity recognition system has been created, and the structure of the set of rules defined on our system. Section 5 describes design and procedure of the experiments carried out. In Section 6, the analysis and the evaluation of these experiments are depicted. Section 7 reviews how the research questions have been accomplished. Finally, we conclude this paper in Section 8.

II. RELATED WORK

The HRI field as a distinct branch of academic activity first emerged in the mid 1990s, although the robot's behaviour and their consequences for humans have been studied in several fields. Goodrich et al. [10] present a survey of current and historical research into HRI. The field is focused on studying robotic systems that interact directly or indirectly with humans. The understanding, evaluation and appropriate design of these systems should facilitate satisfying and naturalistic social interaction between robots and humans. For an assistive robot to be useful for its user at home, the ability to recognize and respond to human activities is essential.

As we mentioned in Section I, the integration of tools such as human activity recognition systems is a first step towards the target of naturalistic interaction between users and robots. In the field of Smart Houses, we can find a huge variety of activity recognition studies, but relatively few are oriented towards robot companions and take into account the need for a reduction of time invested by users in the development of such systems, or the realistic experiments conditions pointed out by Logan et al. [11]. Our ARS has been designed and evaluated based on these principles.

In the literature, two main categories can be found regarding activity recognition systems [8]. The first is based on visual sensors, e.g. camera-based systems to monitor behaviours and changes in the environment [12] [13]. This approach combines computer vision techniques and pattern recognition. The second category is based on sensor networks for monitoring activities in Smart Houses. It can be subdivided into data-centric, logical or semantic approaches. These approaches typically require extensive data collection with potential users of such systems. The data is then analysed using data mining or machine learning techniques to build activity models, which can then form the basis for activity recognition systems. The knowledge-driven, rule-based system approach that we describe in this article belongs to this second category. Similar approaches can be found in the literature [14] [8], but in their evaluation stages participants were told to perform certain activities following a sequence of actions. The approach here presented is capable of recognizing user activities without restricting the way in which users perform those. The use of the non-intrusive sensor network installed helped us create the natural environment the we were looking for. In our view, wearable sensors could affect users' comfort and seem particularly problematic for elderly people.

Other issues have been taken into consideration as well. The system was designed to be easy to move and install in other similar environments without the necessity of specialized knowledge on how these systems work and need to be set up. The rules and sensors are defined in the configuration files (see Section IV-B), followed by a natural language description in order to make the system more understandable. A key advantage of this approach, is that the rules are explicitly represented rather than implicitly represented (e.g. within a Bayesian network [15] [16] or a Hidden Markov Model implementation [17] [6]). This allows us to inspect and manually change or update the rules if needed. As part of the ACCOMPANY project [18], our research in the Robot House will be incrementally developing more complex HRI scenarios for home assistance, so it is important for us to be able to have a system that can be extended and modified easily by non experts, and at the time, keeping the development cost of the system down. We argue that developing a low-cost and resource-efficient system (e.g. the ARS presented), is an important prerequisite for a possible future use in real world applications.

III. Research Questions and Goals

The purpose of this article is to present the development and implementation of the knowledge-driven ARS system and its first validation study. The comparison between the activities recognized by the system, and the actual, observed activities performed by the user during several sessions, will determine the accuracy level of the system and its capacity to be integrated into future HRI studies. The data collected in the first validation study will be used to improve the first set of system parameters and to suggest new features for future versions of the system. In addition, we try to learn about users' behaviour in a natural home situation, and understand how robot companions could behave in such home environments. Our research questions are:

- Q1. Is our ARS generic enough to detect different users' activities without the system being individually trained for the users?
- Q2. Can the ARS achieve an accuracy higher than 80% in the controlled experiments?
- Q3. Can the ARS achieve an accuracy higher than 80% in the uncontrolled experiment?
- Q4. What are the advantages and disadvantages of the ARS presented in this paper?

Figure 1. The UH Robot House layout and sensor arrangement. 59 sensors are available in the house, but only 52 were used and shown here. The two cameras' locations during the experiments are represented in this picture.

The percentages defined in questions 2 and 3, have been set at these values in order to validate the system with an adequate confidence level. This will ensure a reasonable reliability of the environmental information that will be sent to robot companions in future HRI experiments. An accuracy over 80% seems sufficient since robots' behaviour will not solely be based on the information received from the ARS, but supported by the Robot House's system that makes decisions based on further environmental information. Therefore, we expect that this additional information supplied by the ARS will help us improve the robot's awareness of the situation and thus further enhance its abilities when interacting with users in a living environment.

IV. Human Activity Recognition Framework

A. Robot House Sensor Network Description

Two different but complementary commercially available sensor systems, the GEO System and ZigBee Sensor Network, were installed in the Robot House. Both the GEO System and ZigBee Sensor Network have a refresh rate of 1 Hz, which is deemed as adequate to detect user activities.

The GEO System [19] is a real-time energy monitoring system for electrical devices. It is used to detect the activation and deactivation of electrical appliances by the Robot House's users (e.g. opening the refrigerator or boiling water in a kettle). The status of the electrical appliances connected to this system can be queried from the GEO System database.

The ZigBee Sensor Network [20] is used to detect user activity that cannot be detected by the GEO System such as opening of drawers and doors, occupation of chairs and sofa seat places, opening of cold and hot water taps etc. The ZigBee Sensor Network consists of five ZigBee Wireless modules, which are spread across the Robot House. Together they transmit readings from a total of 26 reed contact sensors, 4 temperature sensors and 10 pressure mats to a ZigBee gateway (XBee Gateway X4). The ZigBee gateway forms an interface between ZigBee Sensor Network and the Robot House Ethernet infrastructure, where the ARS resides.

 ISBN: 978-1-61208-250-9

Table I
BEHAVIOUR CODING SCHEME. ACTIVITIES CONSIDERED FOR THE ACTIVITY RECOGNIZER.

Code	Behaviour	Description
ut	Using Toaster	The time that this appliance is switched on
uk	Using Kettle	The time that this appliance is switched on
pf	Preparing Food	The user is in the kitchen preparing some food
pcd	Preparing Cold Drink	The user is having some cold beverage
phd	Preparing Hot Drink	The user is preparing either tea or coffee
co	Computer ON	The time that this appliance is switched on
uc	Using Computer	The user is sitting in the dining area and using the computer
sd	Sitting Dining Area	The user is sitting in the dining area
lt	Laying Table	The user prepares the table before having meal
md	Having Meal Dining Area	The user is sitting in the dining area and having mea
std	Spare Time Dining Area	The user is reading a book or newspaper in the dining area
wt	Watching TV	The user is sitting in the living room and watching the television
t	TV ON	The time that this appliance is switched on
slr	Sitting Living Room	The user is sitting in the living room
stl	Spare Time Living Room	The user is reading a book or newspaper in the living room
ml	Having Meal Living Room	The robot reminds the user about some medicine
ct	Cleaning Table	The user finish the meal and tidy up all the objects used

B. Implementation

The ARS was developed in Java with a local MySQL database for logging purposes. The software consists of the following four different modules:

- ZigBee module. Manages sensor data from ZigBee Sensory Network.
- GeoSystem module. Pulls sensor data from GEO System Database.
- Activity Recognizer module. Analyses the sensory data retrieved from the ZigBee Module and GEO System Module to determine the user's activity.
- User Interface module. Displays and records the detected user's activities and sensory information to a local database (MySQL) and external log files.

The ARS has been tested on both Linux and Windows systems, with a local MySQL database for data logging purposes. The system is configured by using two XML files. The first configuration file contains the representation of the Robot House sensor network (i.e. mapping sensors' IDs to their symbolic names), and the second configuration file defines the semantic rules used by the ARS to detect user's activities in the Robot House (see Section IV-C). These rules were set based on an initial set of trials and the common-sense knowledge which activities of daily living (ADL's) are based on [8]. In future work, the parameters could be refined based on the information gathered after this study. We have to consider that this first experiment is part of the learning process that we have to follow to achieve our final research goals.

Two issues have to be pointed out in regards to the system. Firstly, the ARS is intended to trigger and present an identification at the starting point of the activities studied (see Table I). We consider that the beginning of each activity is the suitable moment at which robot companions should interact with users to offer their help. Secondly, the possibility of migrating the system to other similar environments has been considered during the development process, so that the editing, redefinition or adaptation of these two configuration files would be sufficient to run the system in these new environments.

C. Rule Definition Example

In this section, we show briefly how the ADL's rules have been defined following common-sense knowledge which make the system understandable to any researcher using it. We studied a variety of activities that will be useful in assistive robotics scenarios in future stages (see Table I). These activities can be described as the combination of sensors activated in the environment, and previously performed activities, namely context-activities. Thus, the system manages two different kinds of activities. Low-level activities are those that are detectable by a single fixed sensor (e.g. the user sitting on the sofa). High-level activities are those that can only be detected by utilising a combination of different sensors, or a combination of different sensors and low-level activities detected. Based on that, each rule is defined using the following tags:

- Duration: The maximum time the activity remains activated in the system. Some activities, e.g. *Using Computer Dining Area* and *Sitting Living Room* (described below), do not consider this tag as they are deactivated based on their associated context-activities or associated sensors' status values.
- Location: The location where the activity is performed.
- Context: Set of activities that has to be fulfilled before the activity is activated. Some activities, e.g. *Sitting Living Room*, do not have any context-activity associated with them. *Interval*: Time window in which the context-activity is relevant for the detection of the activity. *Status*: The required context-activity's state for the activation of the activity.
- Threshold (Sensors' attribute): Minimum value necessary to consider the activity as activated. It is based on the accumulated weight of the sensors triggered.
- Sensors: Each of the sensors involved directly in this activity. They have a *Status*, *NotLatching* (True: The sensor's weight will be only added to the accumulated weight while it remains on, otherwise, its weight is subtracted from the accumulated weight; False: the sensor's weight is added to the accumulated weight once it is on regardless of its later state), and *Weight* fields. Some activities, e.g. *Using Computer Dining Area*, do not have any sensors associated with them.

We can see below the examples rule *Using Computer Dining Area* and *Sitting Living Room*. More examples are available from the author on request:

```
<Activity Name="Using_Computer_Dining_Area">
    <Duration>Nil</Duration>
    <Location>Dining_Area</Location>
    <Contexts>
        <Context Interval="0" Status="activated">
         Sitting_Dining_Area</Context>
        <Context Interval="0" Status="activated">
         Computer_ON</Context>
    </Contexts>
    <Sensors Threshold="0.0"></Sensors>
</Activity>
<Activity Name="Sitting_Living_Room">
    <Duration>Nil</Duration>
    <Location>Living_Room</Location>
    <Contexts></Contexts>
    <Sensors Threshold="0.50">
     <Sensor Status="on" NotLatching="true" Weight="50">
        Sofa seatplace 0</Sensor>
     <Sensor Status="on" NotLatching="true" Weight="50">
        Sofa seatplace 1</Sensor>
    /Sensors>
</Activity>
```

In the first example, *Using Computer Dining Area*, the activity depends on *Sitting Dining Area* and *Computer On*, but no sensors are associated with the activity recognition. For this reason, *Duration* and *Threshold* tags are not considered for this activity, as the activity will be activated only when both context-activities are activated. In the second example, the activity is associated with

Table II
THE OBSERVER XT FORMATTED OUTPUT (LEFT SIDE) AND THE ACTIVITY RECOGNIZER'S EVENT LOGS (RIGHT SIDE). THIS DATA REPRESENTATION HELPED US ANALYSE THE RESULTS AND FIND BEHAVIOUR PATTERNS THAT WILL BE CONSIDERED IN FUTURE WORKS.

Observation	Time_Relative_hms	Duration_sf	Behavior	Event_Type		System	System Events	Time	Time Relative	Delay (seconds)
								08:21:35		
User-001-S2	00:00:00	60.74	Preparing_Cold_Drink	State start		Yes	Preparing_Cold_Drink	08:22:18	00:00:43	00:00:43
User-001-S2	00:00:05	299.88	Preparing_Food	State start		Yes	Preparing_Food	08:21:39	00:00:04	00:00:01
User-001-S2	00:00:21	75.04	Using_Toaster	State start		Yes	Using_Toaster	08:21:58	00:00:23	00:00:02
User-001-S2	00:01:00	0	Preparing_Cold_Drink	State stop						
User-001-S2	00:01:28	16.96	Laying_Table	State start		Yes	Laying_Table	08:23:01	00:01:26	00:00:02
User-001-S2	00:01:36	0	Using_Toaster	State stop						
User-001-S2	00:01:45	0	Laying_Table	State stop						
User-001-S2	00:02:01	50.06	Using_Toaster	State start		Yes	Using_Toaster	08:23:36	00:02:01	00:00:00
User-001-S2	00:02:11	41.32	Sitting_Dining_Area	State start		Yes	Sitting_Dining_Area	08:23:45	00:02:10	00:00:01
						Extra	Having_Meal_Dining_Area	08:23:45	00:02:10	00:02:10

certain sensors, whose *NotLatching* field makes their activation compulsory to keep the activity activated as well. Therefore, *Duration* is not considered for this activity, since the deactivation of the associated sensors will deactivate the activity.

V. EXPERIMENTAL DESIGN AND PROCEDURE

A validation study was conducted by the Adaptive Systems Research Group at University of Hertfordshire in May 2012 to measure the accuracy of the framework previously explained. The Robot House provides a naturalistic and ecologically acceptable environment to carry out studies into ADL's. The main aim was to measure the accuracy of the system in both controlled and uncontrolled scenarios and collect data for future studies. A sample of 14 adults, unaffiliated with the ongoing research, and aged between 23 and 54 was recruited from students and staff of the University of Hertfordshire. All the subjects first completed a consent form, in which they were informed about the voluntary nature of the experiments, before they performed a two-day experiment, one session per day. Each session lasted approximately 20 minutes.

A. Experimental Setup

The experiments took place in the Robot House in which ARS were installed and configured. All the experiments were recorded on video and audio using two different cameras (see Figure 1) rather than relying on self-reporting. One camera covered the dining area and living room, and the other covered the kitchen. Those were the only rooms where the participants performed the experiments. The cupboards were labelled to make the participants aware of every object's location and create a more natural environment in the sense of knowing where things are located, as they would feel in their own houses. However, users got used to the Robot House facilities after the introductory session as will be explained in the next section.

The ARS generated two different log files for each participant, one per session. The first file stored information on all the sensors activated and deactivated during the experiment, as well as the decision-making process that the activity recognition algorithm was doing in real time. The second file represents the raw sensory data received from the system during the experiment. These raw data can be used to simulate users living in the Robot House in future experimental scenarios in which robots will be included.

B. Experimental Procedure

The experiments were led by the researcher, who introduced and explained the procedure and the house's facilities to each subject. This section took approximately 10 minutes and was only provided for the first session. After this introductory part, during the first (controlled) session the participants were led by the researcher for 20 minutes, while they were asked to perform a number of specific common ADL's using the Robot House's facilities in the way in which they felt most comfortable with. Thus, they were told what activity to perform, but not how to perform it. In the second (uncontrolled) session, we told the participants to spent around 20 minutes simulating 'living' in the house. They were asked to perform whichever activity (based on the facilities shown during the introductory session) they wished during this period of time. Consequently, we exposed the system to two different situations, controlled and uncontrolled, which would help us measure the system's accuracy and analyse human behaviour at a home environment and discover details omitted in the system, respectively. After each session, the participants were asked to complete a questionnaire. They rated the scenarios and the activities in which they were involved. Basic demographic information of each participant was collected in this questionnaire as well. Note, the order of the conditions was not counterbalanced, since the goal of the study was not to compare the two conditions. Also, it seemed important to first expose participants to the controlled condition which helped them to prepare themselves for the uncontrolled condition.

VI. ANALYSIS AND EVALUATION

A. Behaviour coding

Relatively little work (e.g Logan et al [11]) has combined behaviour coding with user activities in Smart Houses. However, many examples of different data annotation studies can be found in the field of HRI, e.g. [21], and Psychology, e.g. [22]. The coding of the video data of the participants activities helped us analyse each session and identify the important events which we were interested in. The Observer XT software supplied by Noldus Information Technology [23] is a commercial software package used for coding, analysis and presentation of observational data.

The first author of this article was the first coder of all the video material. Additionally, following conventions of behaviour coding, a second coder carried out the same process with 10% of the analysed videos in order to perform the reliability test. The Observer XT and the coding scheme shown in Table I were used by both coders, who were asked to familiarize themselves with this coding scheme before the annotation process. They were told to code activities in which users interacted with some of the sensors installed in the Robot House, in order to generate the sequence of activities that each user had been performed. The outcomes were exported to an Excel files in order to be compared to the events generated by our ARS during the analysis stage.

1) Inter-rater Reliability Test: The Kappa Statistic [24] was used to determine the level of agreement between the two different annotations carried out by the two coders. The annotations were paired in Observer XT, and the kappa value was generated automatically for both sessions. The time windows for the reliability analysis was defined as one second. The kappa value for the combined analysis was 0.75, with overall agreement of 76%. This result represents a good agreement rate for both annotations [25].

B. Data Analysis

A final Excel file was built based on the event lists created using Observer XT and the events generated by our ARS (see Table II). The left side of the table represents the events exported from the software. On the other side, the activities recorded by the system were written down together with their starting time. In this way, the results were shown clearly, and allowed to distinguish 'recognized', 'missed', or 'extra-recognized' activities more easily. The last category represents those activities that fulfilled all the sensor's activations required but they were not performed by the user as evident in the video data. In future experiments, the interaction between the robot companion and the user will help us clarify the real status of these kinds of activities. Moreover, additional tools to support our ARS will be integrated into the Robot House's system during the ACCOMPANY project [18].

A total of 14 participants and two sessions per participants have been considered for the data analysis. We will explain each session separately. The system performance was calculated in terms of precision, recall and accuracy [26] (see Figure 2).

$$Precision = \frac{tp}{tp + fp} \qquad Recall = \frac{tp}{tp + fn}$$

$$Accuracy = \frac{tp}{tp + fp + fn}$$

Figure 2. Precision, recall and accuracy formulas. (tp = true positives or 'recognized', fn = false negatives or 'wrongly recognized' and fp = false positives or 'extra-recognized').

1) Session 1 (controlled): We have to remember that in this scenario the user was lead by the researcher, as we described in Section V. A total of 240 events were coded in all the experiments carried out in this session. The average number of performed activities per user was 17. We got 239 correctly recognized activities, 1 missed activity, and 37 extra-recognized activities were triggered. We obtained a precision of 86,59%, a recall of 99,58% and an accuracy of 86,28%. We found some delay in the recognition of the most complex activities, i.e. those activities involving a major number of different sensors (e.g. preparing food or preparing a beverage). The rest of the activities were recognized with an average delay of two seconds, which is reasonably fast, taking into account the operating system frequency 1Hz.

2) Session 2 (uncontrolled): In the second session, good overall results were achieved too, even taking into account the openness of the scenario which we exposed our system to. A total of 216 events were coded in the experiments carried out during this session. The average number of performed activities per user was 15. We got 200 correctly recognized activities, 16 missed activities and 23 extra-activities were triggered. We obtained a precision of 89,69%, a recall of 92,59% and an accuracy of 83,68%. As stated before, some delay were found on the most complex activities. In Figure 3, we represent these averages delays per activity (e.g. Preparing Hot Drink was recognized with a delay of 35 seconds). The rest of activities were recognized with a similar average delay than in Session 1. The data collected along this experiment will help us understand human behaviour at home and improve our system.

VII. Discussion

The results presented above allow us to answer the research questions presented in Section III. The approach followed has demonstrated the possibility of creating a low-cost, resource-efficient ARS and presenting it to real users without the necessity of previous training. This is directly related to the reduction of time spent by participants in HRI studies as it was mentioned in Section I. The accuracy in both controlled and uncontrolled sessions exceeded the 80% threshold previously defined in our research questions, which was considered as adequate for the kind of study. Some of the advantages presented by this approach are the creation of a non-restricted and naturalistic system that allows users to behave as they would in their own houses. As we mentioned, in other approaches experiments were typically much more constrained. The use of hidden, non-intrusive sensors installed around the Robot House helped us create this natural environment, as we focussed on avoiding wearable sensors that could make users uncomfortable. In addition, the system can be easily migrated and setting in a similar environment without the necessity of specialized knowledge. The rules and sensors were defined using a natural language in order to make the system more understandable.

On the other hand, the system does have some disadvantages. Firstly, the types of sensors currently used do not allow to determine accurately where the user is located in the house. Therefore, the recognition of activities for two or more users simultaneously cannot be detected directly, as the system is not able to match activities with users. An extra tool (e.g. a camera recognition system) may solve the problem, so that it will be considered in future work. However, this will increase cost and complexity of the system and involve privacy issues. Secondly, the semantic rules used by the ARS were defined based on common-sense knowledge of how a person would carry out the ADL's. A module to modified these initial definitions as the user interact with the system will be considered in future stages of our research.

Once the ARS has been integrated into the Robot House system, we will be able to create much richer scenarios in which robot companions will be aware of users' activities. This will allow us to adapt robots' behaviour to their needs in each situation,

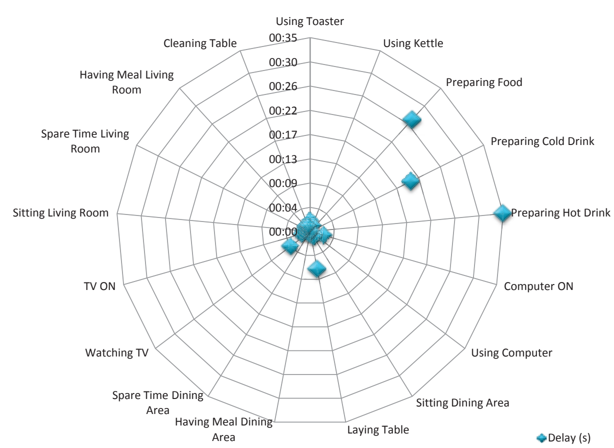


Figure 3. Overall delay per activity in the uncontrolled scenario.

 ISBN: 978-1-61208-250-9

and increase robot companions' autonomy to make decisions. A variety of challenging studies will be targeted in future stages of our research.

VIII. Conclusion and Future Work

We have presented the development and validation of a knowledge-driven rule system to identify user activities in home scenarios. We tried to build a low-cost, resource-efficient and easily understandable and re-configurable system that is accurate enough to detect a set of ADL's. This approach was evaluated empirically by means of the studies carried out in the Robot House. The experimental environment allowed participants to behave in a similar way that they would in their own homes, as it was reported in the questionnaires. Although the participants did not belong to our target user group, i.e. elderly people, we claim that, due to the general design of our system, the results can be generalized, and if necessary, can be easily adapted to this users group. In future work, the adaptation to individual users and their specific life styles and routines may also be considered. The results achieved fulfil our expectations and answer fully the research questions defined in Section III. These findings motivate us to progress towards our final research target of designing context-aware companion robots for home environments. It can be concluded that the developed ARS could be integrated into future experiments of our research.

Acknowledgement

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 287624, the ACCOMPANY project [18].

References

[1] K. Dautenhahn. Socially Intelligent Robots: Dimensions of Human–Robot Interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):679–704, 2007.

[2] K. L. Koay, D. S. Syrdal, M. L. Walters, and K. Dautenhahn. Living with Robots: Investigating the Habituation Effect in Participants' Preferences During a Longitudinal Human–Robot Interaction Study. ROMAN'07, pages 564–569. IEEE, 2007.

[3] K. L. Koay, D. S. Syrdal, M. L. Walters, and K. Dautenhahn. Five Weeks in the Robot House - Exploratory Human–Robot Interaction Trials in a Domestic Setting. In *Proceedings of the 2009 Second International Conferences on Advances in Computer-Human Interactions*, ACHI'09, pages 219–226, Washington, 2009.

[4] M. L. Walters, K. Dautenhahn, R. Te Boekhorst, K. L. Koay, D. S. Syrdal, and C. L. Nehaniv. An Empirical Framework for Human-Robot Proxemics. AISB2009, pages 144–149, 2009.

[5] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A Survey of Socially Interactive Robots. *Robotics and Autonomous Systems*, 42(3-4):143 – 166, 2003.

[6] T. van Kasteren, A. Noulas, G. Englebienne, and B. Kröse. Accurate Activity Recognition in a Home Setting. In *Proceedings of the 10th international conference on Ubiquitous computing*, UbiComp '08, pages 1–9, New York, NY, USA, 2008.

[7] Z. Z. Bien, H. E. Lee, J.-H. Do, Y. H. Kim, K.-H. Park, and S.-E. Yang. Intelligent Interaction for Human-Friendly Service Robot in Smart House Environment. *International Journal of Computational Intelligence Systems*, 1(1):77–94, 2008.

[8] L. Chen, C. D. Nugent, and H. Wang. A Knowledge–Driven Approach to Activity Recognition in Smart Homes. *IEEE Transactions on Knowledge and Data Engineering*, 24:961–974, 2012.

[9] M. Chan, D. Estve, C. Escriba, and E. Campo. A Review of Smart Homes-Present State and Future Challenges. *Computer Methods and Programs in Biomedicine*, 91(1):55 – 81, 2008.

[10] M. A. Goodrich and A. C. Schultz. Human-Robot Interaction: a Survey. *Found. Trends Hum.-Comput. Interact.*, 1(3):203–275, January 2007.

[11] B. Logan, J. Healey, M. Philipose, E. M. Tapia, and S. Intille. A Long–Term Evaluation of Sensing Modalities for Activity Recognition. In *Proceedings of the 9th international conference on Ubiquitous computing*, pages 483–500. Springer-Verlag, 2007.

[12] J. Hoey and J. J. Little. Value–Directed Human Behavior Analysis from Video Using Partially Observable Markov Decision Processes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(7):1118–1132, July 2007.

[13] T. B. Moeslund, A. Hilton, and V. Krüger. A Survey of Advances in Vision-Based Human Motion Capture and Analysis. *Comput. Vis. Image Underst.*, 104(2):90–126, November 2006.

[14] S. Holger, M. Becker, and M. Riedl. Rule-Based Activity Recognition Framework: Challenges, Technique and Learning. In *PervasiveHealth*, pages 1–7. IEEE, 2009.

[15] E. Tapia, S. Intille, and K. Larson. Activity Recognition in the Home Using Simple and Ubiquitous Sensors. In Alois Ferscha and Friedemann Mattern, editors, *Pervasive Computing*, volume 3001 of *Lecture Notes in Computer Science*, pages 158–175. Springer Berlin / Heidelberg, 2004.

[16] L. Bao and S. Intille. Activity recognition from user–annotated acceleration data. In Alois Ferscha and Friedemann Mattern, editors, *Pervasive Computing*, volume 3001 of *Lecture Notes in Computer Science*, pages 1–17. Springer Berlin, 2004.

[17] D. Sanchez, M. Tentori, and J. Favela. Activity Recognition for the Smart Hospital. *Intelligent Systems, IEEE*, 23(2):50 –57, March 2008.

[18] ACCOMPANY: Acceptable robotiCs COMPanions for AgeiNg Years. Project No. 287624. http://www.accompanyproject.eu/, October 2011. [Online; accessed 02.12.2012].

[19] GEO: Green Energy Options. http://www.greenenergyoptions.co.uk/. [Online; accessed 02.12.2012].

[20] S. Farahani. *ZigBee Wireless Networks and Transceivers*. Newnes, Newton, MA, USA, 2008.

[21] K. L. Koay, K. Dautenhahn, S. N. Woods, and M. L. Walters. Empirical Results from Using a Comfort Level Device in Human-Robot Interaction Studies. In *Proceedings of International Conference on Human Robot Interaction*, pages 194–201, 2006.

[22] J. L. Flenthrope and N. C. Brady. Relationships Between Early Gestures and Later Language in Children With Fragile X Syndrome. *Am J Speech Lang Pathol*, 19(2):135–142, 2010.

[23] Noldus Information Technology. The Observer XT Software. http://www.noldus.com/human-behavior-research/products/the-observer-xt/. [Online; accessed 02.12.2012].

[24] J. Sim and C. C. Wright. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, 85(3):257–268, 2005.

[25] R. Bakeman and J. M. Gottman. *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge University Press, 1997.

[26] D. L. Olson and D. Delen. *Advanced Data Mining Techniques*. Springer Publishing Company, Incorporated, 1st edition, 2008.

 ISBN: 978-1-61208-250-9

Comparison of Simultaneous Measurement of Lens Accommodation and Convergence in Viewing Natural and Stereoscopic Visual Target

Tomoki Shiomi
Nagoya University
Nagoya, Japan
shiomi.tomoki@a.mbox.nagoya-u.ac.jp

Keita Uemoto
Nagoya University
Nagoya, Japan
uemoto.keita@d.mbox.nagoya-u.ac.jp

Takehito Kojima
Nagoya University
Nagoya, Japan
kojima.takehito@f.mbox.nagoya-u.ac.jp

Masaru Miyao
Nagoya University
Nagoya, Japan
miyao@nagoya-u.ac.jp

***Abstract*— Recent advances have been made in 3D technology. However, the influence of stereoscopic vision on human sight remains insufficiently understood. The public has come to understand that lens accommodation and convergence are mismatched during stereoscopic vision, and this is the main reason for visual fatigue caused while viewing 3D images. The aim in this study is to compare the fixation distance of accommodation and convergence in viewing real objects and 3D video clips. Real objects and 3D video clips perform the same movements; therefore, we measured accommodation and convergence in subjects who watched both. From the result of this experiment, we found that no discrepancy exists in viewing either 3D video clips or real objects. Therefore, we argue that the symptoms in viewing stereoscopic vision may not be due to the discrepancy between lens accommodation and convergence.**

Keywords-component: accommodation, convergence, simultaneous measurement, stereoscopic vision

I. INTRODUCTION

Recently stereoscopic images have been used in various ways. In spite of this increase in 3D products, and the many studies that have been done on stereoscopic vision, the influence of stereoscopic vision on human visual function remains insufficiently understood. When viewing stereoscopic images, people sometimes feel visual fatigue, 3D sickness, or other discomfort [1].

Investigations of the influence of stereoscopic vision on the human body are essential in order to ensure the safety of viewing virtual 3-dimensional objects People often report symptoms such as eye fatigue and 3D sickness when continuously viewing 3-dimensional images. However, such problems are unreported with so-called natural vision. One of the reasons often given for these symptoms is that lens accommodation and convergence are inconsistent during the viewing of 3D images [1, 2, 3].

Accommodation is a reaction that occurs due to the differences of refractive power by changing the curvature of the lens with the action of the *musculus ciliaris* of the eye along with the elasticity of the lens. The result is that the retina focuses on an image of the external world. Convergence is a movement where both eyes rotate internally, functioning to concentrate the eyes on one point to the front (Fig. 1).

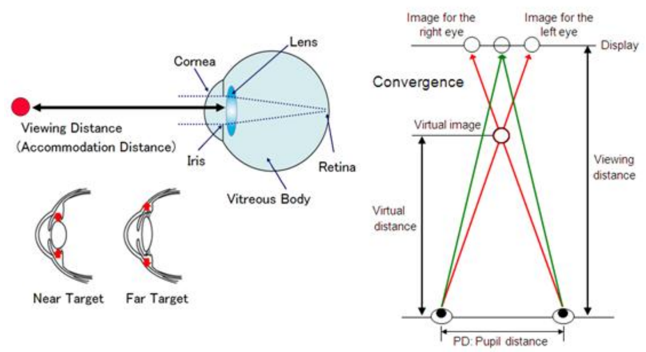


Figure 1. Lens accommodation and convergence

The relationship between accommodation and convergence is one factor that enables humans to see one object with both eyes. Convergence occurs when an image is captured differently with both eyes (parallax). At the same time, focusing on an object is achieved by accommodation. The main method of presenting 3-dimensional images is through the manipulation of the viewers mechanism of binocular vision, and many improvements have been made in this technology.

We suggest that a discrepancy between accommodation and convergence does not exist even when viewing in stereoscopic vision. Our previous study obtained results that indicate that the supposed inconsistency between accommodation and convergence does not occur [4]. In this present study, we performed a more detailed investigation confirming the non-existance of this discrepancy. In section 2, we explained how to measure accommodation and convergence simultaneously, and we showed the result of our experiment in section 3. Then, we discussed our experiment in section 4. Finally, we stated our conclusion and future works in section 5.

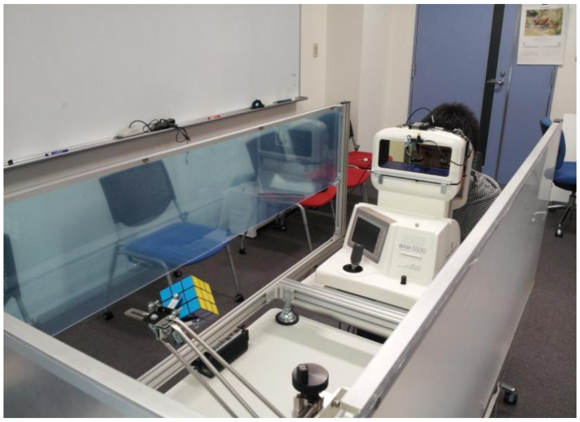


Figure 2. The scene of the measurement

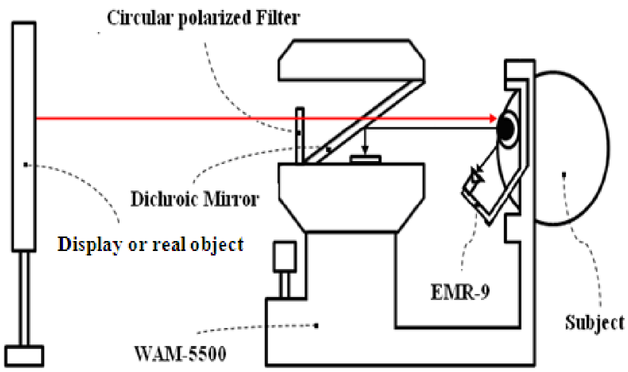


Figure 3. The overview of the measurement process

II. METHOD

We used an original machine developed by combining WAM-5500® and EMR-9® to perform the simultaneous measurements of accommodation and convergence. The experiment was conducted with the help of seven subjects (male and female)

Subjects gazed in binocular vision at a real object in natural vision (a Rubik's cube) and then at a virtual object of 3D video clips presented in front of them (Fig. 2). We measured their lens accommodation and convergence (Fig. 3). The objects viewed by the subjects in natural and stereoscopic vision showed exactly the same motion, and there were three kinds of movements of these objects (Fig. 4).

(1) The objects of natural and stereoscopic vision moved forward and backward at a range from 0.5 to 1m with a cycle of 10 seconds. It was repeated four cycles per single measurement.

(2) The second movement was the same motion as in movement one, but the time of a single cycle of movement was 2.5 seconds.

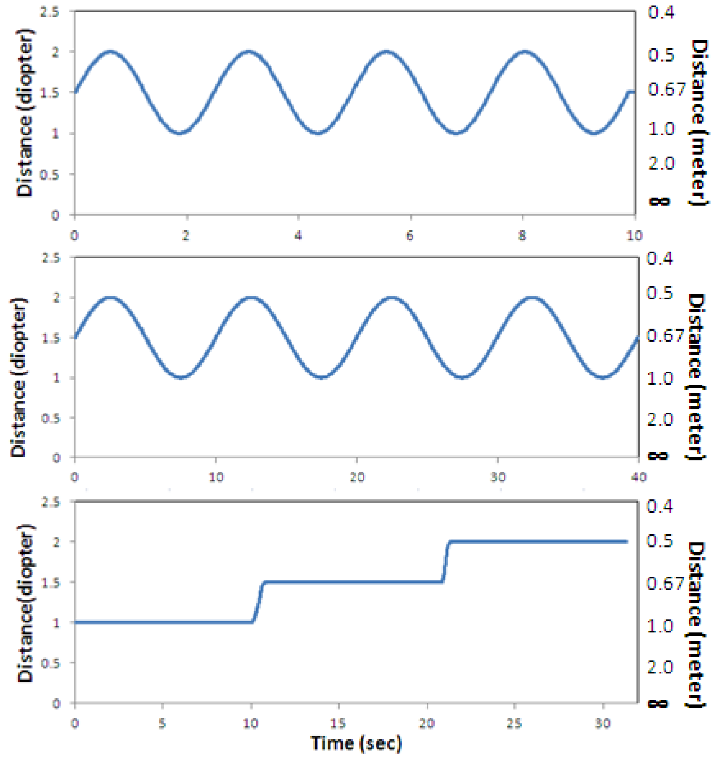


Figure 4. The movement of the object in natural and stereoscopic vision. (1) First was a cycle of 10 seconds. (2) Second was a cycle of 2.5 seconds. (3) Third was step motion, the object stopped at 1D, 1.5D, and 2D for each 10seconds.

(3) The object in this movement approached the subject. Initially, the position of the object was 1m from the subject. The object moved forward to the subject and stopped at the position of 1D, 1.5D, 2D for each 10 seconds (D represents diopter). A"diopter" is the refractive index of the eye lens, which is an index of accommodation power. It would be as follows 0D stands for infinity, 1D stands for 1 m, and 2D stands for 0.5m.

The measurements of the objects in both natural and stereoscopic vision were taken three times per one movement. The illuminance in this experiment was 103 lx.

III. RESULT

The measurements for all subjects showed roughly similar tendencies. Figs. 5 and 6 showed the results of movement 1, which were the moving objects in both natural and stereoscopic vision with a cycle of 10 seconds. The subjects in Fig. 5 and Fig. 6 were different.

Then, Figs. 7 and 8 showed the result of movement 2.

In all these figures, "accommodation" stands for the focal length of lens accommodation, while "convergence" stands for the convergent focal length, and "object" stands for the location of the real object in natural vision or the position of virtual image in stereoscopic vision.

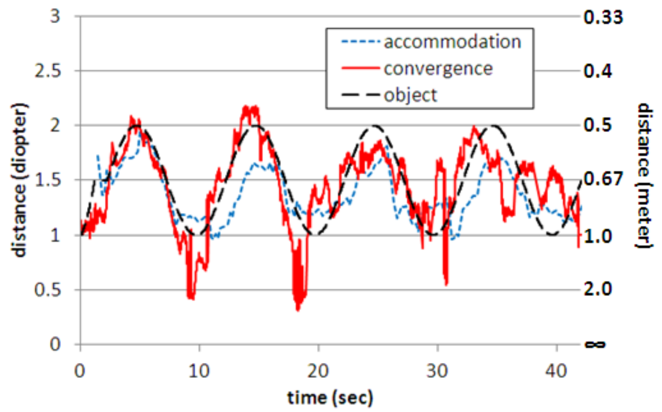


Figure 5. The result of natural vision (a cycle of 10 seconds)

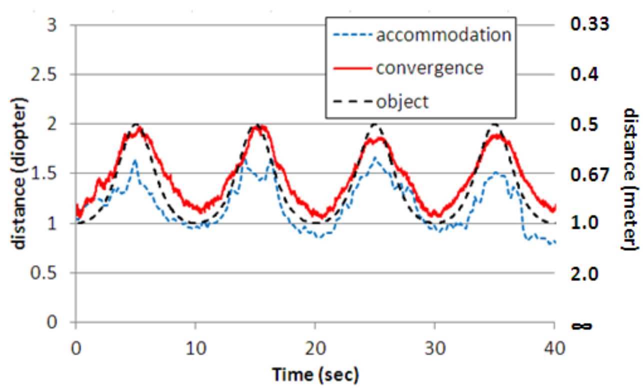


Figure 6. The result of stereoscopic vision (a cycle of 10 seconds)

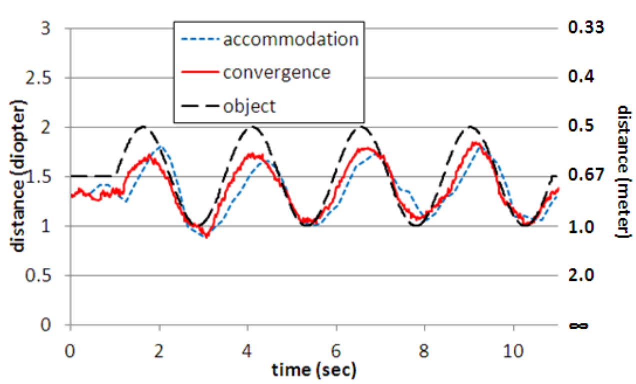


Figure 7. The result of natural vision (a cycle of 2.5 seconds)

All figures show that the accommodation and convergence of subjects changed in agreement. In Figs. 5 and 6, the change in the diopter value occurred within a cycle of about ten seconds and those in Figs. 7 and 8 did within a cycle of about 2.5 seconds. Moreover, the value nearly agreed with the distance from the subject to the position of the real object or virtual image. In the case of movement 3, lens accommodation and convergence also approximately agreed with the position of the virtual images.

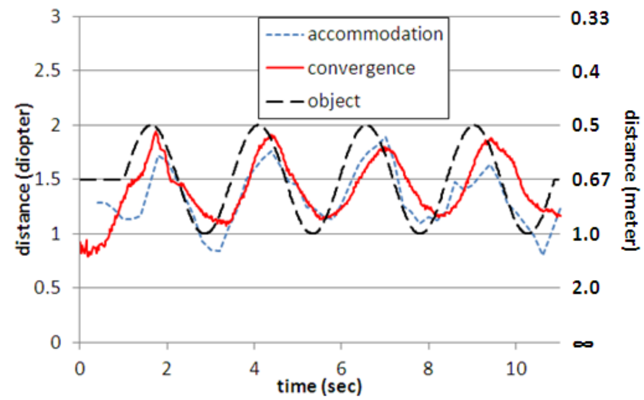


Figure 8. The result of stereoscopic vision (a cycle of 2.5 seconds)

IV. DISCUSSION

According to Hoffman [2] and Ukai & Howarth [3], lens accommodation in viewing 3D images would be fixed at the position of the display. However, or experiment found no mismatch in the accommodation-vergence.

In our previous study, we also reported the results of simultaneous measurement of lens accommodation and convergence while subjects viewed objects in stereoscopic vision, and the inconsistency between accommodation and convergence did not occur [5]. This study simultaneously measured accommodation and convergence in viewing 3D video clips of three movements, and the discrepancy was unconfirmed as in viewing real object. Therefore, we found that subjects watching 3D do not show any discrepancy between accommodation and convergence.

Subjects should be seeing blurred images if lens accommodation focuses on the virtual image position while a stereoscopic image project outwards. Subjects focusing on a nearer position rather than the display may be experiencing the condition in which humans look at a position beyond the farthest point of the object as in myopia.

Smith [6] showed that the relationship between the refractive error and visual acuity is linear. The visual acuity of subjects in Smith's experiment did not decrease much. Therefore, the distance from an emerging object in our experiment may not have been a problem and was correctly viewed by subjects.

Meanwhile, Patterson [7] reported that there should be a problem in only a near-eye display and that the accommodation-vergence mismatch likely would not occur under most stereoscopic display viewing conditions because of the depth of field.

Patterson [7] and Wang [8] also found that the depth of field was large, and they stated that the average total depth of focus was on the order of 1.0 diopter. Based on this value, the range of total depth of field would be from a distance of about 0.1m in front of a fixed point to about 0.17m behind the fixed point of 0.5m. For a fixed distance of 1m, the total depth of field would be from a distance of about 0.33m in front of the point to about 1.0m behind the point. For a fixed distance of 2 m, the total depth of field would be from about 1m in front of the point to an infinite distance behind the fixed point.

They also reported that the depth of field was affected in various ways by the pupil diameter and resolution.

Some researchers found that pupil diameter will be slightly over 6 mm for a luminance level of 0.03cd/m2 and near to 2 mm for a luminance level of 300cd/m2. For each millimeter of decrease in pupil size, the depth of field increases by about 0.12 diopters [7, 9].

Therefore, the value of accommodation can be in the range of the depth of field in our experiment.

In the future research, we plan further investigations concerning the influence of age, pupil diameter, the illuminance of the experimental environment, and the luminance of visual targets.

V. CONCLUSION

In this experiment, we simultaneously measured accommodation and convergence of subjects viewing real objects and 3D video clips. The video clips showed exactly the same motion as with the real objects. We did not confirm the existence of discrepancy between lens accommodation and convergence. Therefore, we believe it is inconclusive that symptoms such as eye fatigue and 3D sickness are caused by this discrepancy and be the result of other factors. We plan to perform further investigations including the change of pupil diameter in viewing stereoscopic vision, luminance level and the difference of the 3D presenting method in order to further comprehend the mechanism of stereoscopic vision.

ACKNOWLEDGMENT

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (B) Number 24300046.

REFERENCES

[1] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, “Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review.” J. Imaging Sci. Technol. Vol. 53, 03, 2009

[2] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks “Vergence-accommodation conflicts hinder visual performance and cause visual fatigue.” Journal of Vision, Vol.8, No.3, 33, pp1-30, 2008

[3] K. Ukai and P. A. Howarth “Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observation.” Displays, Vol.29, 2, pp106-116, 2008

[4] M. Miyao, S. Ishihara, S. Saito, T. Kondo, H. Sakakibara and H. Toyoshima “Visual accommodation and subject performance during a stereographic object task using liquid crystal shutters.” Ergonomics, Vol.39, 11, pp1294-1309, 1996

[5] H. Hori, et al., “Comparison of accommodation and convergence by simultaneous measurements during 2D and 3D vision gaze.” FORMA Special Issue., 2011

[6] G. Smith, “Relation between Spherical Refractive Error and Visual Acuity.” Optometry Vision Science, Vol. 68, No.8, pp591-598, 1991

[7] R. Patterson, “Human factors of stereo displays: An Update.” Journal of SID, 17, 12, pp987-996, 2009

[8] B. Wang and K. J. Ciuffreda, “Depth of focus of the human eye: Theory and clinical applications.” Surv. Ophthal. 51, 75, 2006

[9] K. N. Ogle and J. T. Schwartz, “Depth of focus of the human eye.” J. Opt. Soc. Am. 49, pp273–280, 1959

 ISBN: 978-1-61208-250-9

Study of a FCMAC ANN for Implementation in the Modeling of an Active Control Transtibial Prosthesis

Andrade, J. A. A., Diniz, E. H., Borges, K. C.
Figueiredo, J. F., Silva, R. C.
Faculty of Gama/Gama - FGA
University of Brasília – UnB at Gama
Brasília-DF, Brazil
e-mail: andrade.jaa.unb@gmail.com

Brasil, L. M.
Program of Post-Graduation in Biomedical Engineering
University of Brasília-UnB at Gama
Brasília-DF, Brazil
e-mail: lmbrasil@gmail.com

***Abstract*— This article presents some topics that have been developed by the authors on the development of a model of posture and behavior control of a Active Transtibial Prosthesis, i.e., applied to individuals with amputation below the knee. It is intended to use a Neuro-Fuzzy ANN as the basis of this control. Specifically, we intend to use an FCMAC ANN Neuro-Fuzzy type. Such ANN has the ability of memorizing a region of operation and allow that for similar entries to those known in memory, know outputs may be generated. It is intended to show an early version of this work whose application was the modeling of the inverse kinematics of a leg in the sagittal plane.**

Keywords-component; Active Transtibial Prosthesis, Control, Artificial Intelligence, ANN Neuro-Fuzzy.

I. INTRODUCTION

The search to improve the quality of life for people who have suffered amputation is one of the goals of a branch of engineering called the Rehabilitation Engineering [1]. Using tools and techniques developed in large research centers, it is possible to obtain models of prostheses and orthoses that may allow greater comfort to its user, particularly for those with lower limb amputations.

Modeling the control of prosthesis or any device used in human rehabilitation allows for better interaction with its user and can ensure greater comfort in performing movements [2]. In particular, for prostheses applied to the lower limbs, their behavior during walking is a factor that can facilitate or hamper the movement of the user. In this context, some tools from studies in the area of Artificial Intelligence can assist in the implementation of the control of these devices for rehabilitation. Of these tools, we highlight the use of fuzzy systems and Artificial Neural Networks (ANNs) combined in a system called Neuro-Fuzzy ANNs. The fuzzy systems deal with linguistic sets that allow managing inaccuracies in a way to generate the output more plausible by rules [3]. In the other hand the ANNs seek to simulate the connectionist reasoning using processing units with nervous system biological inspirations. They are able to learn, generalize and classify the elements in the universe they have been trained. The combination of these two systems can leverage the best of each [3].

In this work, we intend to show the results of some studies done about a Neuro-Fuzzy called FCMAC. This ANN has great potential for use in real-time control and allows the memorization of the environment in which it will be applied in order to generate the closest know output [4]. The FCMAC ANN is used as a basis for implementing a posture and behavior control of an active prosthesis for transtibial amputees and, as a first version, the results generated in this study were applied to generate a modeling of inverse kinematics of a leg in the sagittal plane.

II. CHOICE OF A NEURO-FUZZY ANN

Among the various models of Neuro-Fuzzy ANNs, we highlight the ANFIS ANN (Adaptive Network-based Fuzzy Inference Sistem) and FCMAC (*Fuzzy-Cerebellar Model Articulation Controller*) ANN. In previous works [11, 12] we sought to determine which of these two types would be used in the implementation of the control of the active transtibial prosthesis. The following will show some characteristics about these two Neuro-Fuzzy ANNs.

A. ANN ANFIS

The ANFIS is one of the more widespread Neuro-Fuzzy ANNs and represent a type of RBF (Radial Basis Function) ANN in a system of type Fuzzy TSK (Takagi, Sugeno and Kang) [3]. It relies on the use of ANNs and Fuzzy systems whose output has not defuzzification processes. Figure 1 shows the layout and operation of such an ANN.

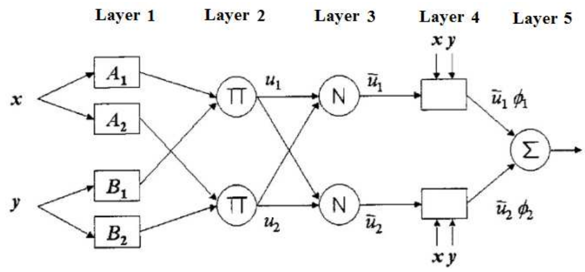


Figure 1. Model of Neuro-*Fuzzy* ANN of the *ANFIS* type [3].

In Figure 1, the first layer represents the fuzzification of the inputs x and y. Layer 2 is responsible for the activation of the outputs of layer 1. Layer 3 is responsible for normalizing the output of layer 2. Layer 4 presents parametric equations on the basis of inputs from the ANN and layer 5 is the sum of parametric equations whose biggest influence on the output comes from the parametric equations whose activations in layers 2 and 3 are larger.

This ANN presents two modes of construction [3]. The first is called Concatenated Model, whose entries are a concatenation of relevance vectors associated with each variable. The second is called the Combined Model and, unlike the previous, its entries are formed by combinations of fuzzy sets of the variables. In this last, to develop a complete system, all possible combinations for these assemblies must be considered.

B. The FCMAC ANN

The FCMAC ANN is an improvement of an architecture proposed by [5], called CMAC. The CMAC (*Cerebellar Model Articulation Controller*) ANN is a simplified model inspired in the cerebral cortex of mammals that operates through associative memories that relate the inputs of the ANN with appropriate outputs. A feature that should be mentioned is that the CMAC ANNs have a large number of receptive fields with finite boundaries [3, 4]. This means that it operates in a mapping by limited intervals of operation. The ANN initially proposed by [5] implements functions of the type shown in (1) and (2).

$$f : S \rightarrow A \tag{1}$$

$$g : A \rightarrow Y \tag{2}$$

In Equation 1, *S* is the space of input from ANN, *A* is the set of CMAC memories and *Y* is the output. The mapping of CMAC memories is done using functions of Boolean activation. The outputs of these activations are combined using an AND Boolean operator to enable or disable your memories. The memories that are activated are weighted by weights and summed to generate the output of the ANN. However, the use of activation functions of Boolean type in entries, makes some CMAC memories completely connected or disconnected. This makes its output response to present discontinuities, even for well-behaved inputs [6].

An alternative proposed by [7] to solve the problem of discontinuities in the output of the ANN was the adoption of activation functions of fuzzy logic. These functions work with concepts of pertinence of elements and allow use as a activation one degree trigger that allows certain memory locations that are not completely disconnected from the output pattern and are usually Gaussian or B-splaines [3, 6]. Thus some memory locations may contribute fully, partially or not contribute to the output, making that this way it can be continuous and smooth. The structure of a Fuzzy-CMAC ANN (or FCMAC) can be seen in Figure 2, where it notices that it is very similar to the CMAC ANN, differentiated only by the activation functions in the entries.

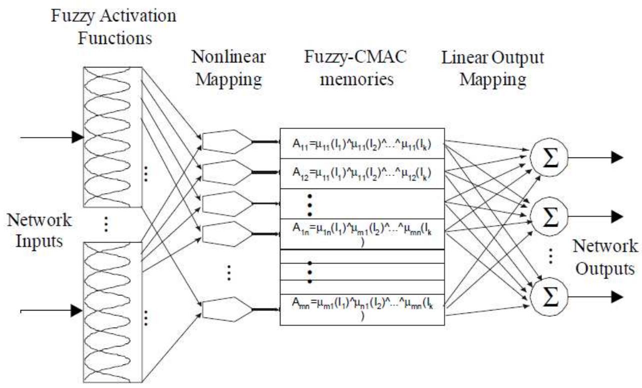


Figure 2. FCMAC ANN structure. Adapted from [6, 4].

C. Justification for the Selection of FCMAC ANN

Regarding ANFIS ANN, it appears that it seeks to preserve the properties of generalization of ANNs. Moreover, the knowledge of an expert can serve to start the ANN parameters so that during your training, it can give better results by already be closer to the desired setting. However, with regard to its constitution, it must be verified its limitations for problems not linearly separable, that is, with similar characteristics but belonging to different sets [8]. This is because in the case of the combined model, its application in project is subject to the number of entries in the ANN, since the number of combinations grows exponentially with the number of inputs, besides being more difficult to construct. Now with regard to the concatenated model, its construction becomes easier, however, the ANN becomes less robust.

Regarding the Fuzzy-CMAC ANN, a feature that should be mentioned is that their receptive fields act as memories of Fuzzy-CMAC ANN, by enabling or disabling certain ANN connections. Thus, only a few receptive fields will predominantly contribute to the output of the ANN while others will not contribute effectively. Thanks to this, it is possible to conduct a local training ANN only fields that contribute to output and allowing some information stored by the ANN are not lost. Moreover, the problem of discontinuities solved by the use of continuous functions of fuzzy logic, may apply a method of training using a gradient method [3, 6].

Thus, it appears that the Fuzzy-CMAC ANN is presented as a robust technique that can map the characteristics of the human gait. Memorize certain features of the walk can allow the prosthesis to have a consistent stance that can relate the user's intentions with the environment [12]. Given its characteristics, the aim is to use a Fuzzy-CMAC ANN as the basis to perform the modeling control of the active prosthesis for transtibial amputees.

 ISBN: 978-1-61208-250-9

III. MATERIAIS E MÉTODOS

A. Operation of FCMAC ANN

Likewise the CMAC ANN, the FCMAC also operates at finite boundaries, i.e., inputs are limited to their ranges of operation. At the entrance of ANN, μ values are calculated which are the degrees of pertinence of the input values in the range of operation. Thus the output of the activation functions are calculated in (3).

$$\mu_{ij} = \exp\left(-\frac{1}{2}\left(\frac{x_i - c_{ij}}{\sigma_{ij}}\right)^2\right) \tag{3}$$

In Equation 3, μ is the value of the pertinence of input *x*, partitions of the interval *c* and *σ* are the center and support of the membership functions, *i* ranges from 1 to n inputs and *j* ranges from 1 to k number of membership functions. Then the combinations of the membership functions of the inputs are combined with each other as the operator E. For FCMAC ANN are used t-norm operators as minimum or algebraic product [9]. For this work, we used the algebraic product given by (4).

$$a_m = \prod \mu_{ij}(x_j) \tag{4}$$

The index *m* in (4) ranges from 1 to k^n number of rules or combinations of inputs, and *a* is an associative cell related to the FCMAC memory of this multiplicand, and related to the insert input. The associative cells activated by input values that are closer to the values entered in the ANN showed a higher value. With that they influence the output more than those with a lower value. However, all cells contributed to the output of the ANN that in this study was adopted as the (5) [6].

$$Y = \frac{\sum_{m=1}^{N} w_m a_m}{\sum_{m=1}^{N} a_m} \tag{5}$$

where *N* is the number of memories in the ANN.

B. Learning of FCMAC ANN

For the correction of the weights of ANN, one can use an equation similar to the correction of the weights of CMAC ANN. It is given by (6). It is also possible to adjust the Gaussian activation functions of the input layer. This adjustment is made using the backpropagation algorithm and the variables *c* (center) are modified and *σ* of the Gaussian functions. This adjustment improves performance at the ANN output [6].

$$\Delta w_{ij} = \eta (Yd - Y) \frac{a_j}{\sum_{m=1}^{N} a_m} \tag{6}$$

C. Reduction in the number of inputs combinations

A problem that occurs with this ANN is that the total number of possible rules, or combinations of functions grows exponentially in a relation between the number of entries *n* and the number of membership functions *k* as in (7). This is referred to as curse of dimensionality [3], since the number of entries of ANN may limit the application of the designer.

$$C = k^n \tag{7}$$

To reduce this problem, [6] used a limited number of combinations and adjusted the membership functions in the inputs so that the generated combinations could map the inputs in the best possible way. For this he treated the membership functions or activation using an algorithm of clustering or categorization that allowed these functions to initialize. The clustering algorithms used to start the membership functions was the Fuzzy C-Means. With this technique it is possible to calculate the centers of Gaussians activation functions so that they are well distributed in the input space. This allowed the number of combinations of inputs to grow linearly with the number of activation functions for each input, as shown in (8).

$$C = k \tag{8}$$

In (7) and (8), C is the number of possible combinations of inputs FCMAC. The Fuzzy C-Means algorithm is one of the algorithms most utilized for clustering and as well as other algorithms of the type, aims to minimize a function of the type presented in (9), which represents a cost function based on the distance of each point in the input assembly, in the position of each cluster and in the relevance of each point to the cluster [10].

$$J = \sum_{i=1}^{c} \sum_{j=1}^{P} \mu_{ij}{}^{m} d^2(x_j, v_i) \tag{9}$$

In (9), x_j is a sample of a point vector *P*, *c* is the number of desired clusters, *d* is the distance from x_j to the center v_i and μ_{ij} is the value of degree of relevance of the sample x_j to the center v_i, *m* is called fuzzy factor which typically is equal to 2 (the bigger *m is*, it is said the cluster is more *fuzzy* [10]) and *J* is called cost function.

D. Implementation of a FCMAC ANN

To implement a FCMAC ANN, it was necessary to list what are the elements that characterize an ANN of this type. Thus, some elements were listed, and as a result, was generated a data structure that could represent ANN as illustrated in Figure 3. To implement FCMAC ANN, it was used the computer álgebra ambient Matlab ® because it allows greater ease in the development of algorithms.

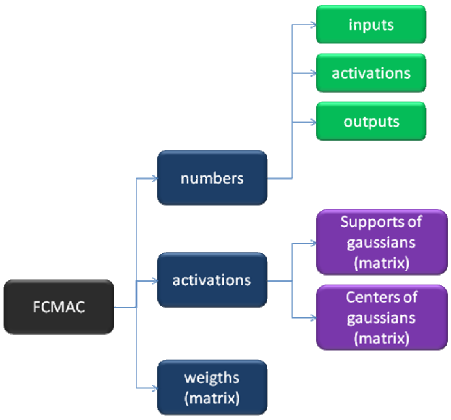


Figure 3. Data structure that represents the FCMAC ANN.

In Figure 3, the attribute numbers stores values corresponding to dimensions of ANN, such as number of entrances, activation functions and numbers of outputs. The attribute activations stores two matrixes that store the values of centers and supports of the Gaussian functions of the inputs. Finally, the attribute weights keeps the corresponding matrix weights of each output, each initialized with random values between -1 and 1 [6].

E. Neuro-Fuzzy Control for a Active Transtibial Prosthesis

The active transtibial prosthesis has a set of sensors that allow for the acquisition of kinematic and dynamic parameters during operation. As the initial proposal, the control of posture and behavior will not consider the bioelectric signals of the user, ie, will be used only dynamic kinematic parameters acquired by the sensors. For this purpose, it is intended to train a FCMAC ANN in order to bring consistent output based on the parameters obtained from sensors. Thus, based on previous work [11, 12], were collected some parameters that feed the FCMAC ANN:

- Distance d relative to the ground plane;
- Angle α between the sole of the prosthesis to the ground plane;
- Angles β between the sole and the leg prosthesis;
- Force F applied to the prosthesis due to the weight of the user;

With these parameters, it is proposed the control model shown in Figure 4 [13].

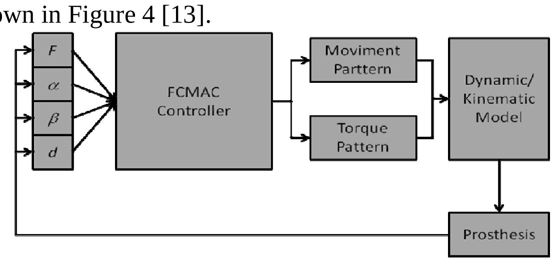


Figure 4. Proposed model to control the prosthesis [13].

In Figure 4, it can be seen that the output of FCMAC ANN should provide at its output the movement to be performed by the prosthesis and how much torque must be applied to a joint. This pattern of torque must be applied at the ankle and depend on the phase of gait. The movement pattern also depends on the phase of gait. These two standards should be adjusted according to the kinematic and dynamic model of the prosthesis so that your posture and behavior can be consistent.

IV. RESULTS

As a first version in the application of Neuro-Fuzzy FCMAC ANN in a control template for a Active Transtibial prosthesis, worked on a model of inverse kinematics of a leg in the sagittal plane. This means that one wishes to obtain the angles which allow the end of the leg may be located in a certain position. For this, it is considered a model with two segments and two degrees of freedom (degree of motion). The ANN which controls the position of the leg has as input parameters a position in which the point on the ankle is in the *xy* plane and an output, it is expected angles of the hip and knee. The structure of ANN used can be seen in Figure 5.

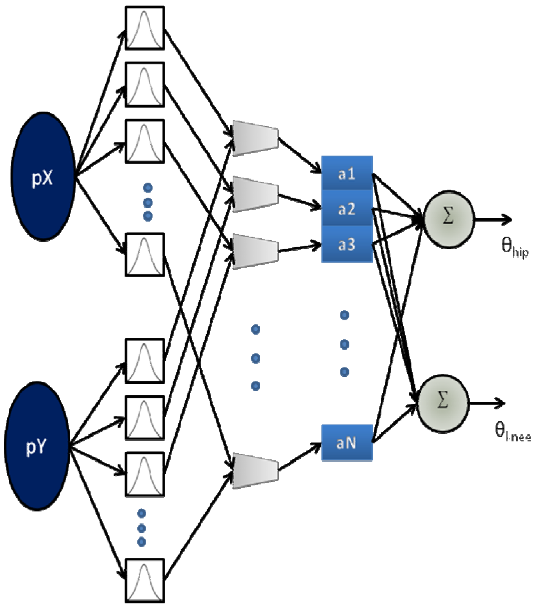


Figure 5. Structure of FCMAC ANN to control the leg where N = 100, that is, the ANN possesses 100 memories.

For training the ANN, it was generated a leg operation space in which its edge could be located. In this operating region, about 400 training points were generate, which were also used to generate a 100 clusters that serve as centers for the Gaussian functions activation inputs of ANN. The region restriction is shown in Figure 6, where two segments of line represents the leg segments and the red points are the clusters generated in this mapping.

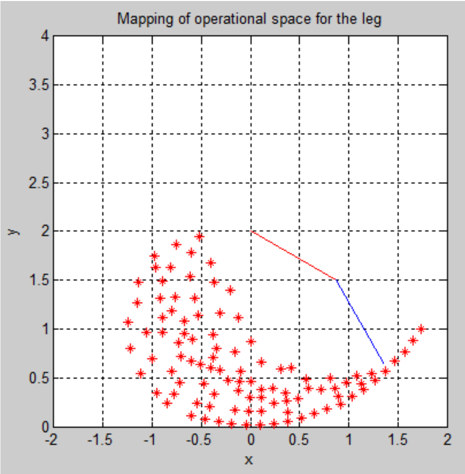


Figure 6. Mapping of the leg operation space. Strictly in this region, the FCMAC ANN can provide the desired output after training.

Established these points, began training a FCMAC ANN with about 100 activation functions on each input. After training, was traced a trajectory parameterized in the *xy* plane and tried to verify if the ANN would be able to perform the inverse kinematics of this trajectory to position the end of the leg in this plan. Figure 7 shows the trajectory traced to the end of the leg and the solution of inverse kinematics for the angles corresponding to this path.

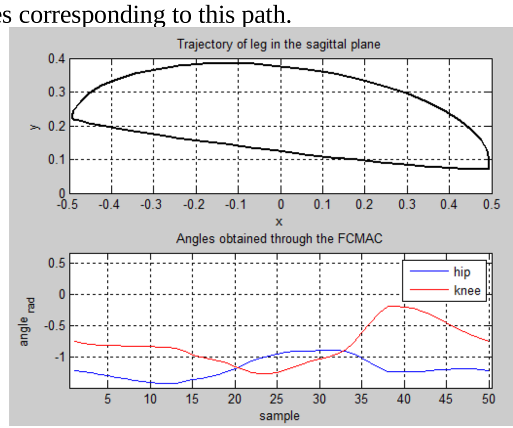


Figure 7. At the top of the figure has the trajectory desired to be approximate by FCMAC ANN. In the lower part there are the corresponding angles to this path

V. Discussion

When using a categorization algorithm, one can find some clusters that are positioned at locations around whose elements have some similarity. By doing so, it was possible to perform a mapping of the leg operation space with a limited amount of clusters equal to ¼ of data that would be used in training. This could allow a drastic reduction in the number of combinations of inputs needed so that the ANN could approximate the trajectory. Thus, for the ANN used, which had 100 functions in the activation of 2 starters, it took only 100 combinations of inputs instead of 1002 = 10000.

The FCMAC ANN training for this application was carried out in 500 times and wished the Mean Square Error (MSE) was 10^{-4}. However, at the end of training epochs, the NDE obtained was 1.6 x10^{-4}. However, ANN performance was satisfactory, because the behavior of the trajectories obtained in simulation was close to expectations as you can see in Figure 8, where they are shown briefly simulation performance of the FCMAC ANN to approximate the trajectory of the leg.

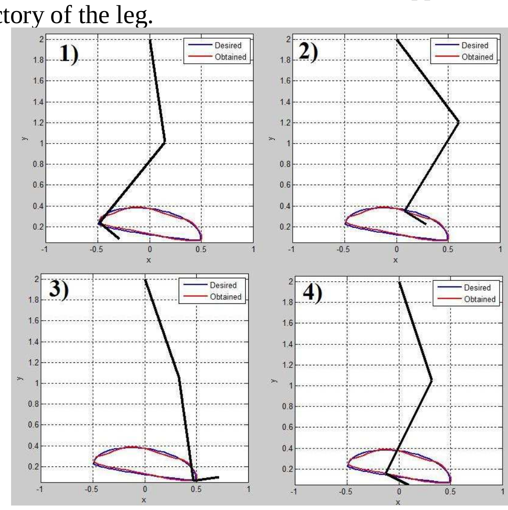


Figure 8. The figure shows a few moments of trained FCMAC simulation. The blue curve represents the desired trajectory and the red represents the trajectory obtained by ANN.

VI. Conclusion

Through this work, we can learn a few points about the capabilities of FCMAC ANN. It was found that in case of using all possible input combinations, their applications on design would be constrained by the number of entries as the number of combinations grows exponentially with the number of inputs and activation functions.

It was seen that the use of a categorization algorithm allows the number of input combinations is dramatically reduced to a linear complexity to the number of activation functions of the inputs. The Fuzzy C-Means algorithm allowed for mapping the operating region of the leg so that could distribute the activation functions of the inputs through this region.

As regards the application done, verified the ability of FCMAC ANN memorizing the operating region and allow the execution of any path in this region even without the known during training. That's why when associating various positions in the area of operation with the angles of the hip and knee for the training data, the ANN memorized these relationships. Thus, when designing a trajectory in this region, it checks the path which points are closer to the region known demand and provide an approximate output.

It should be noted also that the coordinates are entered in the input ANN were operating in the area, leaving the ANN

 ISBN: 978-1-61208-250-9

would be inconsistent. Thus, mechanisms are needed to prevent these boundary conditions occur.

As a next step, we intend to use the data obtained in a simulation environment developed in [11, 12] to train a FCMAC ANN. The aim is also to obtain a model of posture control and behavior of active prosthesis for transtibial amputees.

ACKNOWLEDGMENT

The authors thanks the University of Brasilia (UnB) and the Ministry of Health for the financial support through the Scientific Initiation Scholarship program and another financial suports.

REFERENCES

[1] Cassemiro, E.R., "Methodology for the Development of Biomechanics Devices for the Application in Anthropomorphic Prosthesis", Thesis of Master, FEM/UNICAMP, Campinas, 2002, 89 p. (In Portuguese).

[2] Blaya, A.J., Herr, H., "Adaptative Control of a Variable-Impedance Ankle-Foot Orthosis to Assist Drop-Foot Gait", In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. Vol. 12, p., 24-31, 01 Sept 2004.

[3] Rezend, S.O., Intelligent Systems – Fundamentals and Applications , São Paulo: Manole, 2002 (In Portuguese).

[4] Almeida, P.E.M. ; Simões, M.G. , "Parametric CMAC networks: Fundamentals and applications of a fast convergence neural structure". IEEE Transactions on Industry Applications, New York, NY, v. 39, n.5, 2003, p. 1551-155.

[5] Albus, J. S., "A new approach to manipulator control: The cerebellar model articulation controller". Trans. ASME, J. Dyn. Syst., Meas. Control, v. 97, p. 220-227, Sept 1975.

[6] Reis Jr, S.G., "Thermal Modeling of a Lamination Process of Tubes Using Neural Networks Fuzzy-CMAC", Thesis of Master, DPPG/CEFET-MG, Belo Horizonte, 2005, 132 p. (In Portuguese).

[7] Brown, M., Harrys, C. Neuro-Fuzzy Adaptative Modelling and Control., 1th ed., Prentice-Hall, New York, 1994.

[8] Rich, E., Knight, K. , Artificial Intelligence, USA: Markron, 1992.

[9] Simões, M. G. ; Shaw, I. S., Fuzzy Control and Modeling, 2nd ed. São Paulo: Edgard Bluchert, 2007, 200p. (In Portuguese).

[10] Berni, C. C., "Implementation in Hardware/Firmware of a Virtual Sensor Using Fuzzy Identification Algorithm". Thesis of Master in Engineering – University of São Paulo, São Paulo. 2004. 179p. (In Portuguese).

[11] Silva, R.C., Brasil, L.M., Costa Jr, E.A., Rosa, S.S.R.F., Andrade, J.A,A, Rodrigues, O., "The Development of an Active Prosthesis for Transtibial Amputees: Applied models and Sensorial Fusing", In: Technically Assisted Rehabilitation – TAR Conference 2011, Berlin, p., 1-4, 17-18 mar 2011.

[12] Andrade, J.A.A., Silva, R.C., Araujo, E., Rosa, S.S.R.F., "Contributing to Human Rehabilitation through Study of an Active Prosthetic", In: International Symposium on Applied Electromagnetics and Mechanics - ISEM 2011, Napoli, p., 1-6, 06-09 sept. 2011.

[13] Andrade, J. A. A., Brasil, L. M., et al., "Proposal of Study of Neuro-Fuzzy Networks for the Applications int the Control of a Transtibial Active Prosthesis", In: XXIII Congresso Brasileiro de Engenharia Biomédica (XXIII-CBEB), Porto de Galinhas, oct 2012, 1-3p. (In Portuguese).

The Iterative Design and Evaluation Approach for a Socially-aware Search and Retrieval Application for Digital Archiving

Dimitris Spiliotopoulos
Innovation Lab
Athens Technology Centre
Athens, Greece
d.spiliotopoulos@atc.gr

Ruben Bouwmeester
New Media, Innovation
Deutsche Welle
Bonn, Germany
ruben.bouwmeester@dw.de

Dominik Frey
Documentation and Archives
Suedwestrundfunk
Baden-Baden, Germany
dominik.frey@swr.de

Georgios Kouroupetroglou, Pepi Stavropoulou
Department of Informatics and Telecommunications
University of Athens
Athens, Greece
{koupe; pepis}@di.uoa.gr

***Abstract*—Designing user interfaces involves several iterations for usability design and evaluation as well as incremental functionality integration and testing. This paper reports on the methodological approach for the design and implementation of an application that is used for search and retrieval of socially-aware digital content. It presents the archivist view of professional media organizations and the specific requirements for successful retrieval of content. The content derived from the social media analysis is enormous and appropriate actions need to be taken to avoid irrelevant and/or repeated social information in the displayed results as well as over-information. The archivist feedback reveals the way humans address the social information as presented in the form of metadata along with the archived raw content and how this drives the design of a dedicated search and retrieval application.**

Keywords-search and retrieval user interfaces; social network information; archiving; preservation; user interface design; usability

I. INTRODUCTION

Social Media provide a vast amount of information identifying stories, events, entities that play the crucial role of shaping the community from continuous user involvement [2,3,4]. This work reports on the design and development of a socially-aware search and retrieval application that has the main goal of enabling archivists and researchers to retrieve archived web content based on semantic information derived from social media. Social media categories (SMC) cover almost all existing social media networks that information can be harvested from [1]. This information can be categorized according to type (multimodal: text, video, sound, picture) and role players (agents, users, opinion leaders) and actual semantic meta-data (entities, opinions on entities, etc.). The aim is to design a user interface that uses an engine for semantic analysis taking into account several modalities (plain text, sound, image, video), social media crawling, contextual search fusion and archivist usage requirements. The interface is the means to accessing the vast amount of information that can be archived. This work is ongoing and the findings of the pilot user testing and evaluation provide indications on how the semantic analysis of the social media information can be integrated to the design methodologies for user interfaces resulting in maximization of user experience in terms of social information involvement. The following sections describe the motivation and related work, the initial design considerations and the user evaluation that resulted in updated requirements for the next iteration phase of the design.

II. MOTIVATION

There are several approaches and applications in the recent years for digital archiving. However, these approaches mostly specialize in the archiving process itself without taking into account the actual impact of the archived information to the world. The need for leveraging the wisdom of the crowd for selecting the optimal content for future preservation has been stressed by digital libraries and the broadcasting world [5]. Concrete requirements for reporting the peoples' opinions on important events and associating them with candidate content pages for have been laid out. These stress the important role of the social media for such task. Harvesting the opinions of the people and enabling the retrieval of such social information for the future generations will provide a unique eye in the history and preservation of events. It also validates the selection of the content appropriate for archiving based on the impact of the recorded information that was reported in it. A way of gathering and analyzing such information from the web and building innovative ways to retrieve it was, therefore, a crucial requirement for the above process.

Earlier works treat archives independently of any externally provided social information but allow for manual

 ISBN: 978-1-61208-250-9

or semi-automatic annotation of the relational tags between items [6]. Such applications are in effect annotation interfaces rather than search and retrieval. And for such tasks there were formal models used for the annotation procedure [7]. The approach discussed in this paper goes beyond simple consideration of either an interface for search and retrieval of archived documents or social network dedicated interface. It is a dedicated interface for archived documents that are heavily enriched with social web metadata as analyzed and ingested data derived from social media.

The task before us was to design an innovative interface that would use the semantic information to satisfy complex queries as well as provide semantically enriched views of the archived document information. In order to do that, the user feedback was collected to refine the design considerations.

III. INITIAL DESIGN CONSIDERATIONS

The requirements for the search and retrieval of semantically analysed archived content were considered on the domain, content use and interaction levels. As it can be seen, the social web content integration and the raw data and their relations were a key to this exploration.

A. Domain-specific Requirements

Domain-wise there are certain needs for specific use within organizations. Media organizations have journalists that retrieve selected data on events, topics and entities for articles or documentaries that are under production. Those data need to be as complete as possible and provide a variety of sources and other related articles.

B. High-level Requirements

In the broadcasting world, collections of information targeting a common set of topics and events are called "campaigns". Examples of campaigns span from generalized, large size, long-time types like "The EU financial crisis" to the more focused, short-lived types like "EU Summit 2012" or "Presidential Primaries 2012". The former may span several years, the latter only a few months. Taking the above into consideration, a main functionality for our initial design approach is continuity of data collection and archiving. The scheduled crawled sets and analysis should span the entire campaign both in terms of time as well as in terms of social network content provision. Activity should be monitored and data collected accordingly. The sub-campaigns should correlate to the main campaign that they are supposed to belong to. The users should be able to retrieve information from all relevant archives and continuity of the semantic analysis should be preserved throughout the campaign search and retrieval process.

The database plays an important role in the architecture of such application, as it provides the storing, indexing and retrieving functionality for all the data collected and utilized by the archivists for the refinement process and searched upon by the generic users. As such, it is expected to support various types of data, handle updates and accommodate many types of queries.

The required functionality of the database is defined by its interaction with the search and retrieval user interface.

The database is expected to store: a) the original web content fetched by the crawler during for specific campaigns, b) metadata, meaning the extra attributes that are derived from the application aware crawling (e.g. author information, #retweets for a tweet etc.). The database should be able to answer quantitative as well as qualitative related queries posed by the archivists to further guide the filtering and retrieval processes. All end-users must be able to navigate through the stored content by posing queries that concern either the annotated meta-data or the semantic information derived from the analysis of the social content.

C. Non-functional requirements

Usability testing in our case required the study of users preferences towards the social media derived information. Semantic analysis results have enriched the archived content with semantic descriptions and tags that include entities (persons, locations, dates, etc.), events, topics, opinions on them, trending, cultural dynamics and more. Each web document may include several pieces of semantic information that may or may not be useful to the archivist or end user that would search and retrieve. Before proceeding with mockup design, experimentation has taken place in order to derive knowledge on the user perspective. Results from experimentation with fusing and visualizing social content with semantically driven context-sensitive information were derived [8]. Based on them, the mockups have been designed so that the socially aware semantic information would fully complement the search results. The users have indicated that opinions from social networks were considered fundamental to their understanding of the impact of searched web content. Furthermore, the semantic meta-data were integrated to the design of the interaction flow, so that users may facet and filter their search by using several levels of semantic content as well as prominent traditional information such as social network source information.

IV. THE PROTOTYPE

Iterative design process from the start to the implementation of the first prototype was followed from the initial design requirements gathering (made via wire-framing) to the mockup user environment, to the online prototype. The mockups were constructed based on the three types of information from the functional specifications (Search filtering, Core content, Social content) and the non-functional specifications described above. The initial experimentation was on the observation of the user perception of processed content (filters, tag clouds, paths), direct content (item descriptions, authors, dates, type of modality) and social content (opinions, trends, semantics, entities, events).

Based on the above, low fidelity mockups were created for the web retrieval interface. Those were subsequently evaluated for the core functionalities (filtering of results, follow-up search, result visualization) as well as usability (user approach to semantic search, information load, user effort, acceptance). The initial evaluation resulted in the set of specifications laid out by professional users (archivists, broadcaster researchers). This specification was used for the first version of the application prototype.

The prototype itself used real data from the #greekfinancialcrisis (Twitter) as the main social web source as well as crawled web pages using as parameters the entities that were identified from the #greekfinancialcrisis text analysis. The raw content and the semantic metadata were indexed in Solr. Free text search as well as query enabled semantic search comprised the search functionality. The returned results were web resources that matched the search string. The users were able to view the search results in the main content frame. The results had distinct indicators on the opinion and trending values for each web resource. Dynamically created facets were provided for refinement of the results. Each result was, in fact, web page from the raw content archives. The content that could be viewed were not the original web pages themselves but descriptive data taken from them, such as title, initial description text, source, author. That view was enriched with lists of named entities (people, organizations, locations, etc.), events, and opinions derived from the text analysis. Furthermore, data from the most relevant tweets were provided where the users could use to see the following:

- Timeline showing the opinions of the Twitter authors on the entities contained in the web page
- Lists of positive and negative tweets on the entities of the web resource

V. REFINING THE GOALS

The prototype was used for the user evaluation which, itself, had the main purpose of monitoring the usage in order to construct and validate updated user requirements

A. *Application specifications*

A screenshot of the prototype is shown in Figure 1. The points of interest are pointed out by the highlighted numbers and included the following:

- Single semantic search (Fig.1 point 3)
- Advanced search
- Refine search
- Dynamic filtering via faceting (Fig.1 point 4)
- Sorting (by modality, by source SMC) (Fig.1 point 5, 6)
- Item viewing properties and derived functionalities:
  - o Modality views (all, text, image, video, sound)
  - o Social Media related information per item (current opinion, trending, latest tweets, source SMC)
  - o Text/image analysis related information (entities, events)
  - o Tag cloud support for major entities/events for quick refinement
  - o SMC and text/image analysis items with linked specific functions (preliminary functions include search refinement, new search)

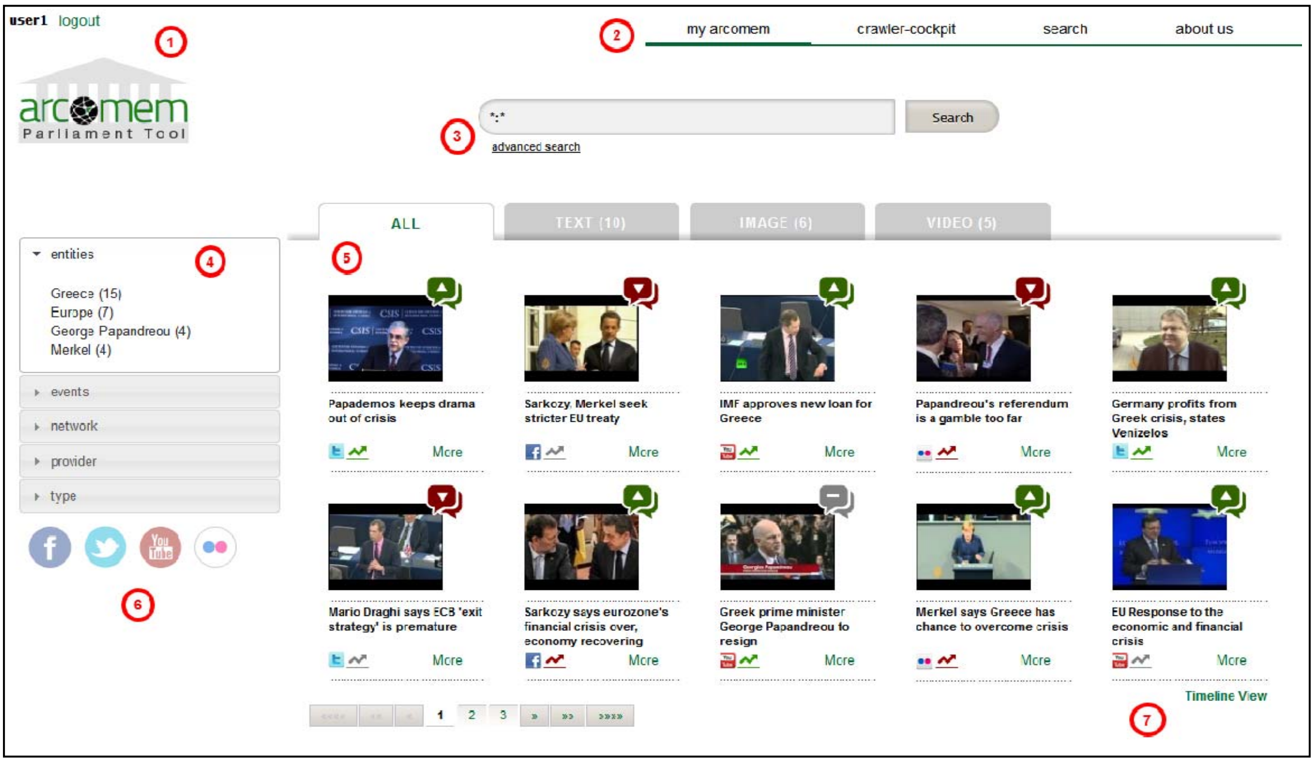


Figure 1. Search results for semantic search within a political domain.

- Generic viewing properties:
  - o Recent searches

- o Recent campaigns
- o Latest news (promotional, user-specific)

Also, as highlighted in Figure 1, main placeholders and associated functions or visual elements we clarified:

- User login and application logo
- Top level menu reserved for high level functions
- Search box (also toggling single and advanced search)
- Dynamic filtering via facets
- Main results page with modality sorting tabs
- SMC filtering (results are filtered by selecting Social Media icons - toggling)
- Timeline view enablement button

The item results page (Fig. 2) provides all retrieved information about an archived web resource. The standard information includes:

- Title
- Details
- Date
- Provider
- Format
- Description

The semantic information includes and is presented to the user as:

- the identified associated entities
- the identified events and topics
- a tag cloud of entities derived from items that include the entities/events of currently web resource.
- The related events of the current identified event (if any)
- the latest tweets that contributed the major entities and opinions for the current web resource
- the major twitter accounts that contributed to the above
- the timeline that shows the positive-negative social network input regarding the associated entities (currently only for Twitter for the first prototype)
- the lists of positive and negative Tweets for the above

B. *Evaluation*

The early prototype that was build based on the previous feedback and specification has been evaluated by the professional end-users in order to refine the goals of the application based on the archivist/broadcaster workflow. So, in that respect, usability was measured qualitatively rather than quantitatively and was guided by specific search and retrieval scenarios.

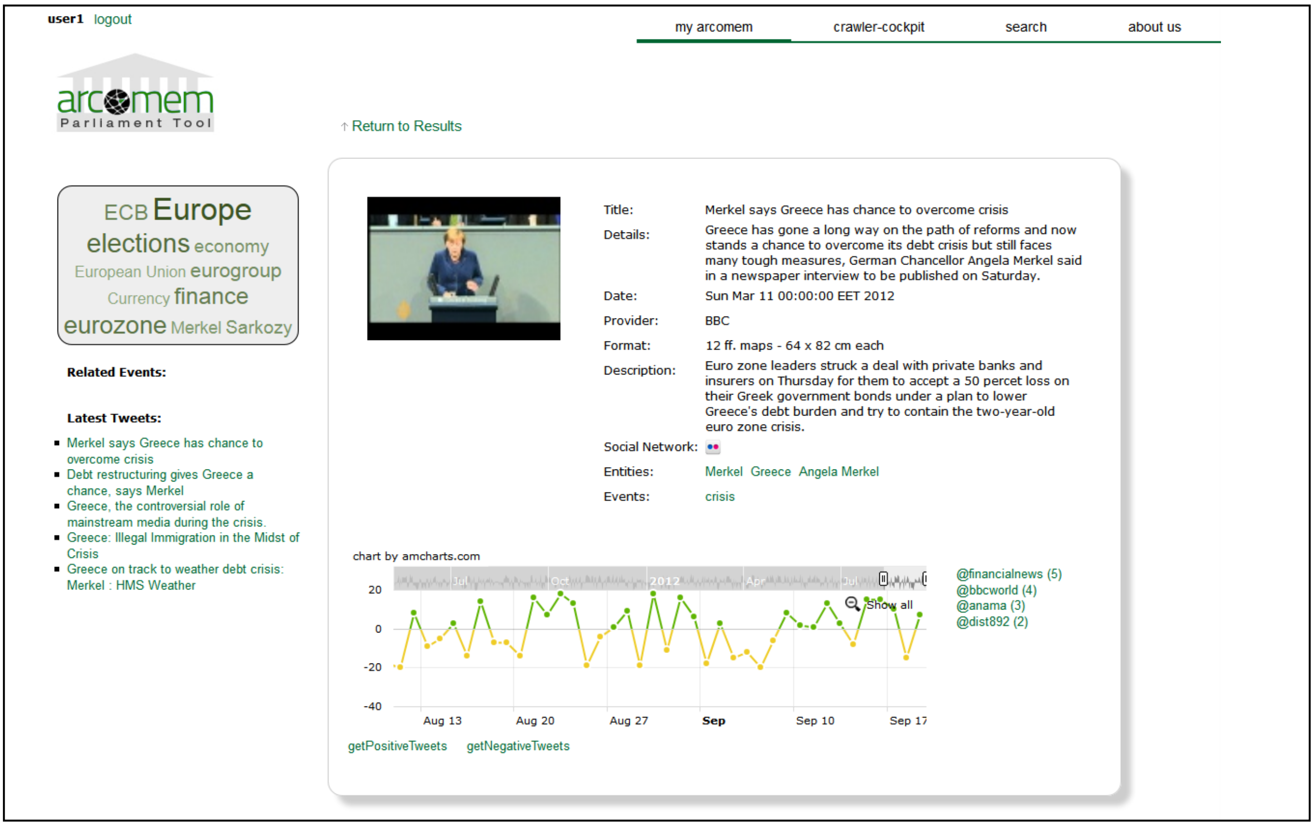


Figure 2. Web object – an archived web page enhanced with semantic information

Evaluation scenarios were created and deployed for novice users. The expert archivists performed free search actions aiming for in-depth social information research on specific content that they knew already. The scope of the evaluation was to enable validation of present functionalities as well as requests for extensions. Another task was to rank the types of semantic information from highly relevant to irrelevant for as many queries as possible.

Those data are fed to the facet ranking algorithm, so that the dynamic facet selection options are optimized. Table 1 provides an overview of the elements of the prototype testing that were evaluated.

TABLE I. EVALUATION OF THE PROTOTYPE APPLICATION

Actions	**Evaluation consolidated overview**		
	Expected results	***Validation***	***Extension***
query	Single and advanced query	Yes	Yes
filter	Facetting/filtering search	Yes	No
single page	Viewing a web page	No	Yes

This evaluation provided an updated set of specifications from the point of usability in terms of system interaction but unavoidably that would be dependent on actual content. The new specification requirements include:

- Ability to search for campaigns (top page functionality) based on campaign tags or description.
- Latest/popular campaigns
- Tag cloud that displays campaign keywords.
- Campaign overview (characteristic entities, events, SMCs, opinions)
- Campaign information (ID, logo, title).
- Information on volume of social/semantic data
- Types of result presentation (list, timeline, map)
- Connection to Wayback machine.

The next iteration will involve the advanced prototype with vast amount of semantic data attached to the archived pages that should be tested both functionally and for usability.

VI. CONCLUSION

This paper reported on the design and testing issues for an application for socially-aware web archiving. Important findings on how archivists and researchers view the social information and how that is integrated to their research workflow have been identified. So far, the social/semantic information is used to guide the search and retrieval process by a large margin. More than 80% of the content used comes from social networks. That shows how important that information is for web archiving and how that information provision can be optimized from a dedicated application interface.

ACKNOWLEDGMENT

The work described here was partially supported by the EU ICT research project ARCOMEM: Archive Communities Memories, www.arcomem.eu, FP7-ICT-270239.

REFERENCES

[1] ARCOMEM: Archive Communities Memories, www.arcomem.eu, FP7-ICT-270239 [retrieved: February, 2013].

[2] K. Church, J. Neumann, M. Cherubini, and N. Oliver, "SocialSearchBrowser: A novel mobile search nd information discover tool", Proc. 14th Int. Conf. Intelligent User Interfaces, ACM, 2010, pp. 101-110.

[3] J. Golberg and M. Wasser, "SocialBrowsing: Integrating social networks and web browsing", Proc. ACM CHI2007 Conf. Human Factors in Computing Systems, April 28 – May 3, 2007, San Jose, USA. pp. 2382-2386.

[4] M. De Choudhury, S. Counts, and M. Czerwinski, "Identifying Relevant Social Media Content: Leveraging Information Diversity and User Cognition", Microsoft Research, ACM, 2010.

[5] G. Schefbeck, D. Spiliotopoulos, and T. Risse, "The Recent Challenge in Web Archiving: Archiving the Social Web", International Council on Archives Congress, 20-24 August 2012, Brisbane, Australia.

[6] M. Agosti and N. Ferro, "A formal model of annotations of digital content", ACM Transactions on Information Systems (TOIS), 26(1), pp. 3:1–3:57.

[7] C. Kohlschütter, F. Abel, and D. Skoutas, "A Novel Interface for Exploring, Annotating and Organizing News and Blogs Articles", Proc. 3rd Annual Workshop on Search in Social Media (SSM 2010), co-located with the ACM WSDM 2010 Conference, 3rd February, 2010, New York City, USA.

[8] D. Spiliotopoulos, E. Tzoannos, P. Stavropoulou, G. Kouroupetroglou, and A. Pino, "Designing user interfaces for social media driven digital preservation and information retrieval", Proc. 13th Int. Conf. on Computers Helping People with Special Needs, 11-13 July, 2012, Linz, Austria, pp. 581-584, doi: 10.1007/978-3-642-31522-0_87

Effect of Agent Embodiment on the Elder User Enjoyment of a Game

Jérémy Wrobel, Ya-Huei Wu, Hélène Kerhervé, Laila Kamali, Anne-Sophie Rigaud
Broca Hospital
APHP & Team Research EA 4468 of Paris Descartes University Paris, France
jeremy.wrobel | yahuei.wu | helene.kerherve | anne-sophie.rigaud @brc.aphp.fr; l.kamali19@gmail.com

Céline Jost, Brigitte Le Pévédic, Dominique Duhaut
Lab-STICC laboratory
University of South Brittany
Vannes, France
and the Robadom Project
celine.jost@gmail.com; brigitte.le-pevedic | dominique.duhaut @univ-ubs.fr

***Abstract*— This paper presents a study that compared the elder user enjoyment of a game of trivia in three conditions: participants playing the game with a laptop PC *vs.* a robot *vs.* a virtual agent. Statistical analysis did not show any significant difference of the three devices on user enjoyment while qualitative analysis revealed a preference for the laptop PC condition, followed by the robot and the virtual agent. The elderly participants were concentrated on the task performance rather on the interaction with systems. They preferred laptop PC condition mainly because there were less interfaces distracting them from performing the task proposed by the game. Further, the robot was preferred to a virtual agent because of its physical presence. Some issues of the experiment design are raised and directions for future research are suggested to gain more insight into the effects of agent embodiment on human-agent interaction.**

Keywords-Embodied agent; human-robot interaction; user enjoyment.

I. INTRODUCTION

In recent years, one area of interests in human-robot interaction studies is to investigate the physical embodiment effects of social agents on their interaction with humans [1, 2]. A physically embodied robot, with both an actual physical shape, embedded sensors and motors and co-located with a human is considered to facilitate better social interaction by prompting human social expectations for proper social interaction than a disembodied or a virtual agent [1, 3, 4]. Several experiments have been conducted, comparing effects of co-located physical robots with remote or telepresent robots and virtual agents engaging humans in different types of tasks. A variety of objective (e.g., task performance) and subjective measures (e.g., enjoyment) have been used to capture these effects (see Table 1 for a review).

Kidd and Breazeal [5] found that a physically embodied robot was considered different than an animated character: it was more engaging, more enjoyable to interact with, and more informative and credible. In another experiment, they showed that participants' perceptions of a physically present robot (co-located with humans) and of a remote one (presented on a television screen) did not differ significantly. As a result, they concluded that what led people to respond differently in the first experiment lies in fact that participants considered both physically present and remote embodied robot as a real and tangible thing, in comparison to the simulated virtual character in the screen. In the studies of Wainer et al. [4, 6], participants rated the physically embodied robot to be more attractive (they spent more time to watch it) and more enjoyable to interact with and more helpful than the virtual robot and the remote robot. Shinozawa et al. [7], showed that a physically embodied robot (three-dimensional body) has more impact on human decision-making when the interaction environment is a three-dimensional space, but has less impact in a two-dimensional space than a virtual on screen robot (two dimensional body). Lee et al. [1] suggested that physical embodiment plays an important role on people's evaluation of social agents even though social agents are not related to any physical function. Furthermore, physical embodiment has an added value for people's social interaction with agents and is an effective means to increase the social presence of an object. Thus, it is an essential aspect of social agents in order to facilitate meaningful social interactions. In the study of Komatsu and Abe [8], most participants accepted the physical robotic agent's invitation to play a game while many neglected the virtual on-screen agent's invitation. The authors suggested that physically embodied robots were considered as more comfortable and believable interactive partners than virtual ones. In the study of Pereira et al. [9], they found that participants who played against the physically embodied agent reported higher enjoyment experiences than those who played with the virtual version of the robot. They suggested that during a computerized chess game, a physical embodied agent elicited a more immersive user experience and a more believable social interaction and led people to believe that they received better system feedback. Hasegawa et al. [10] investigated the impact of embodiment on direction-giving systems by comparing a physically embodied robot, a virtual robot and a Global Positioning System (GPS). The authors found that in the direction-giving systems, both physically embodied robot and virtual robot were more positively rated, in comparison to a simple GPS without any embodiment. However, embodied agents did not allow a better cognitive performances (e.g., retelling of direction-giving), comparing with a simple GPS system.

TABLE I. REVIEW OF STUDIES ON EFFECTS OF AGENT EMBODIMENT ON HUMAN-AGENT INTERACTIONS

Authors	Conditions compared	Interaction tasks	Measures
Kidd & Breazeal, 2004	1. Physically embodied robot *vs*. virtual (simulated) robot *vs*. a human 2. Physically embodied robot *vs*. remote robot (presented on TV screen)	1. The participant responded to spoken requests from the characters, which asked the participant to manipulate colored wooden blocks 2. The desert surviving task and a teaching task	1. Questionnaire assessing enjoyment, informativeness, reliability, fairness, credibility, liking, responsiveness, positivity, looking, involvement 2. Questionnaire assessing sincerity, informativeness, dominance, likeability, reliability, openness, trustworthiness, engagement
Shinozawa et al., 2005	2D task environment + no agent *vs*. 2D task environment + virtual robot *vs*. 2D task environment + physical embodied robot *vs*. 3D task environment + no agent *vs*. 3D task environment + virtual robot *vs*. 3D task environment + physical embodied robot	Color-name selection task	The mean selection ratios of the color names that the agent or robot successfully recommended to each subject (influence of agents on decision-making)
Lee et al., 2006	Physically embodied robot (Sony Aibo) *vs*. virtual on-screen robot	Free interaction with robots	Questionnaire assessing perception of Aibo as a companion, social attraction toward Aibo, enjoyment, public evaluation of Aibo, and social presence
Wainer et al., 2006, 2007	Physically embodied robot *vs*. remote robot (presented on TV screen) *vs*. virtual (simulated) robot	Tower of Hanoi puzzle	Task performance, mean time spent in each condition, questionnaire assessing perception of a social agent's capabilities and the user's enjoyment of the task
Komatsu & Abe, 2008	Physically embodied robot *vs*. virtual on-screen robot	Puzzle video game (picross) with agents	Acceptation of the agent's invitation to play a game, duration of looking at the robotic agent, task performance (number of puzzles solved)
Pereira et al., 2008	Physically embodied robot (iCat) *vs*. virtual on-screen robot	Computerized chess game with agents as co-players	Questionnaire assessing user enjoyment in game
Hasegawa et al., 2010	Physically embodied robot with speaker perspective gesture *vs*. physically embodied robot with listener perspective gesture *vs*. physically embodied robot without gesture *vs*. virtual robot with speaker perspective gesture *vs*. virtual robot with listener perspective gesture *vs*. virtual robot without gesture	Listening to systems for a direction-giving	Performance on a retelling of a direction-giving task, performance on a map task and questionnaire assessing naturalness, presence, engagement, understandability, familiarity, reliability and enjoyment

In the current literature review, we did not find any studies investigating human-agent interactions involving elderly subjects who are often targeted as an important population of end-users of social assistive robots and ambient-assisted living technologies. How do they perceive different kinds of agent embodiment? Is there any added value of the physical embodiment of a robot or a virtual agent compared to a simple PC? In the present work, we designed the following experiment to answer these questions by studying interactions between the elderly and the physically embodied robot, a virtual agent and a simple laptop PC in the situation where they played a game of trivia (e.g., geography, history, literature, etc). We reported our findings on people's perceptions when interacting with the different devices.

II. EXPERIMENT

The design of this study was a within-subjects, repeated measures experiment. Three conditions were set up: subjects interacted with a laptop *vs.* a virtual agent (Greta [11]) *vs.* a physically embodied robot. In each condition, subjects were invited to play a game of trivia StimCards with the system which gave instructions and feedbacks. We compared user enjoyment and engagement within these three conditions.

A. StimCards

StimCards is an interactive card game which is played between a human player and a computing player. The computing player is the game coordinator which asks questions and corrects answers given by the human player. StimCards is composed of:

- A set of game cards with a barcode (a QR code) on the verso. The left side of the Figure 1 shows response items and the right side shows the verso with QR code.
- A camera which can detect QR code and load associated questions.
- A computer screen which displays StimCards GUI and the content of questions. Figure 2 shows StimCards GUI with an example of loaded card.
- An associated input device which allows human player answering questions. Figure 1 shows the input device provided to human player during the experiment.

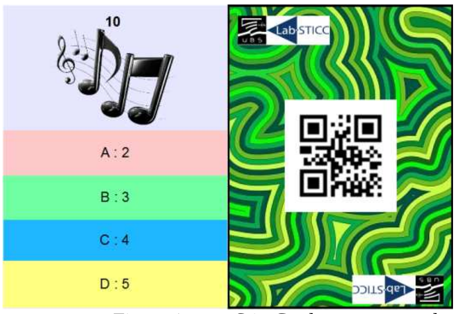


Figure 1. StimCards game example

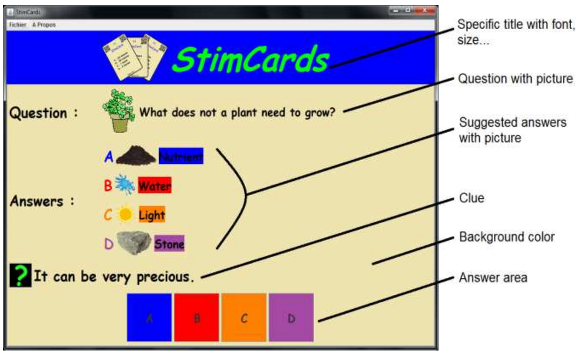


Figure 2. A loaded card in StimCards GUI

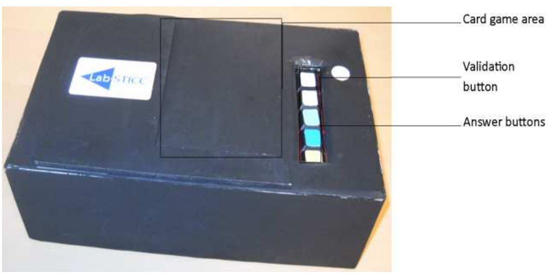


Figure 3. Response box

Each card is associated to a XML file which contains: the question label and associated picture, question type (multiple choice question, open question, etc), a card category (entertainment, sciences, math, etc), GUI background color, font color, a set of clues which can help gamers, a set of suggested answers (text and/or picture) and the correct answer. It is possible to create new questions by changing XML file content or creating new game card associated to new XML file. StimCards is created with MICE [12], a computing modular framework within which a visual programming language creates interaction scenarios allowing digital devices to communicate with each other. Thus, StimCards is configurable in two ways: it is possible to create new cards and to describe the game sequences (interaction scenario).

B. Participants

We recruited nineteen elderly participants with a range in age from 63 to 88 years and with a mean of 75.30 (6.54). There were 3 men and 16 women (Table II). They were contacted by phone from an existing study participant recruitment pool. A consent form was signed by all of the participants before partaking in this experiment. Because of technical problems, data gathered from two participants were excluded for final analysis.

TABLE II BACKGROUND CHARACTERISTICS OF PARTICIPANTS

Gender	n=17	%
Male	2	12%
Female	15	88%
Age		
Mean ± SD	75.68 ± 6.35 years	
Range	63-88 years	
Education		
Mean ± SD	13.2 ± 2.72 years	
Range	9-20 years	
Owner of a computer		
Yes	15	88%
No	2	12%
Frequency of computer use		
everyday - frequently	12	71%
Rarely - never	5	29%
Frequency of computerized game play		
Each day - frequently	1	6%
Rarely - never	16	94%

C. Experimental conditions

Participants played a game of trivia with the following devices (Fig. 4):

 ISBN: 978-1-61208-250-9

a) Laptop PC: subjects were seated in front of a table containing a laptop PC, a webcam, a response box and the playing cards.
b) Physically embodied robot: subjects were seated in front of a table containing a webcam, a response box and the trivia cards. The robot (Robulab of Robosoft) with a screen was placed at the right side of the table.
c) Virtual agent + PC laptop: subjects were seated in front of a table containing a laptop PC, a webcam, a response box and the trivia cards. The virtual agent (Greta) was projected on the wall at the right side of the table.

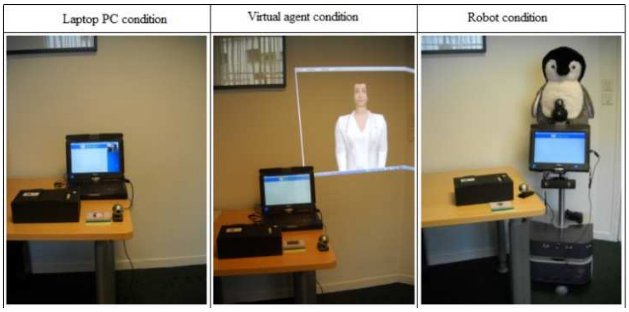


Figure 4. Experimental scene

D. Task

In each condition, participants played a game of trivia. The questions of the game were based on general knowledge, composed of 5 themes (literature, cinema, politic, geography, music). For each question, there were 4 possible answers and only one was correct. Each participant was asked to respond to 10 questions.

We created the following scenario for each condition:

1. Device says "Hello. Press the validation button from the response box when you are ready".
2. Participant player presses the button.
3. Device says "Hello. I will ask you 10 questions. You will use the black response box to put your card and answer the question. Exercises start now."
4. Device says "You can scan a card by placing it in front of the camera".
5. Participant shows a card to the camera.
6. Device says "A card has been detected. Put your card on the response box and press the validation button when you are ready".
7. Participant presses the validation button.
8. Device reads the question while the computer screen displays the question content of StimCards.
9. Participant responds to the question.
10. If participant responds correctly, device says "Congratulation! It is the good answer". If the participant gave the wrong answer, device said "Sorry, it is not a good answer. The good answer is […]".
11. The step 4-10 are repeated nine times. At the end, device says "Exercises are finished. Thank you for your participation".

E. Procedure

Upon arriving at the living lab (Hôpital La Collégiale, Paris), subjects were told the purpose of the experiment. If they agreed to participate, they signed a consent form. In a randomly assigned order, each subject performed the task in three conditions. In each condition, the subject filled out a questionnaire assessing user enjoyment. At the end of the three conditions, they were asked to comment on the three systems and to talk about the system they preferred to interact with.

F. Measures

To evaluate user enjoyment and engagement, we designed a questionnaire that consisted of items based on GameFlow model [13] and United Theory of Acceptance and Use of Technology (UTAUT) [14].

The 5-point Likert questionnaire consisted of 13 items (Table II), measuring four dimensions of GameFlow (*feedback*, *immersion*, *social interaction*, *concentration*) and 2 dimensions of UTAUT (*intention to use*, *perceived enjoyment*). Participants were asked to indicate their level of agreement to the statement following the 5-point response scale anchored by "not agree" and "Totally agree". Moreover, observation and note taking were carried out for qualitative analysis.

TABLE III ITEM OF THE USER ENJOYMENT QUESTIONNAIRE

Domains	Statements
Concentration	• I was caught up in the game. • I stayed focused on the game.
Immersion	• I felt involved in the game. • I forgot about time passing while playing the game.
Social interaction	• I found that interaction with the […] was pleasant. • I appreciated accompanied by the […]. • Playing condition was convivial.
Feedback	• I received immediate feedback on my actions. • I appreciated the feedback given by the […].
Intention to use	• I would recommend this game to people around me. • If the multiplayer mode exists, I would recommend playing this game with my friends.
Perceived enjoyment	• Generally, I enjoyed playing the game. • Alone, I would accept to play the game with the […].

 ISBN: 978-1-61208-250-9

III. RESULTS

A one-way within-subjects ANOVA was conducted to compare user enjoyment in the three conditions. In total, 17 subjects took part of the experiment. The Table IV shows the means, standard deviations and analysis of variance of the global score and sub-scores of user enjoyment in the three conditions. Even though subjects rated higher user enjoyment under the laptop PC condition, compared to the two other conditions, the results of ANOVA did not show any significant differences among the three conditions.

We have performed a qualitative analysis from observation and field notes about how participants interacted with systems and how they considered of them during the experiment. We observed that a majority of the participants were concentrated on the response box and they rarely looked at the screen of the laptop PC, the robot and the virtual agent. In fact, they looked at these devices at the beginning but after a few minutes, they concentrated on manipulating the response box to perform the task.

Further, most of participants reported that they preferred the laptop PC condition to the other conditions because they could be more concentrated on the task. On the other hand, they considered the devices for two other conditions too cumbersome and not easy to use. They did not see any added value of virtual agent and robot when performing the task. For the condition of the robot, some participants appreciated the robot's physical presence compared to the virtual agent. Nevertheless, they judged the robot's head, a stuffed animal, too childish. Furthermore, they found it lacking life-like characteristics. As for the virtual agent + PC laptop condition, only a few subjects appreciated the presence of the virtual agent. Most of the participants criticized it because they found it adynamic and its gaze lacking emotion.

IV. DISCUSSION

This study investigated user enjoyment in a game of trivia in three conditions: laptop PC *vs.* robot *vs.* virtual agent. Although participants rated higher user enjoyment under the laptop PC condition, followed by the robot condition and the virtual agent condition, statistical analysis did not show any significantly difference among the three conditions (Fig. 6). The preference towards the laptop PC condition can be explained by the fact that participants were mainly concentrated on the task performance rather than one "social interaction" with the system. They focused on the response box and rarely looked at the other interfaces except for the PC screen. Some of them even considered that the robot and the virtual agent distracted them from performing the task. Furthermore, several participants reported that there were too many things to look at and they could not pay attention to all interfaces. This result is somehow not surprising as impairments in divided attention and associated executive functions are dominant among the cognitive impairments associated with normal aging [15].

Besides, many subjects conceded that they did not see any added values of a robot or a virtual agent in this kind of task. They said that they did not find it interesting to play with a robot or a virtual agent partly because they lack living characteristics and their appearance is not appealing.

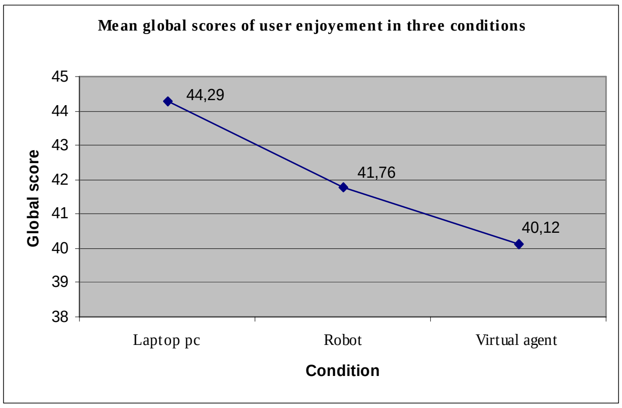


Figure 6. Mean global scores of user enjoyment in three conditions

TABLE IV. MEANS, STANDARD DEVIATIONS AND ANALYSES OF VARIANCE OF THE GLOBAL SCORE AND SUB-SCORES OF USER ENJOYMENT IN THE THREE EXPERIMENTAL CONDITIONS

	Laptop	Robot + laptop pc	Virtual agent + laptop pc	*F*(2,48)	*p*
Global score	44.29 (8.39)	41.76 (9.75)	40.12 (9.27)	1.02	0.37
Feedback	3.44 (0.73)	3.41 (0.76)	3.24 (0.94)	0.32	0.73
Immersion	3.44 (0.83)	3.24 (0.95)	3.08 (0.95)	0.75	0.48
Social interaction	3.67 (0.53)	3.18 (0.96)	2.94 (1.18)	2.70	0.08
Concentration	3.56 (0.63)	3.32 (0.92)	3.47 (0.74)	0.40	0.67
Intention to use	3.15 (1.17)	2.88 (1.33)	2.85 (1.30)	0.28	0.76
Perceived enjoyment	3.18 (0.98)	3.26 (1.00)	3.03 (0.96)	0.25	0.78

 ISBN: 978-1-61208-250-9

On the other hand, our findings also showed that the participants preferred the robot condition compared to the virtual agent condition. According to our qualitative analysis, the advantage of a robot over a virtual agent is that a robot provides a physical presence. For example, some participants said that the robot was tangible and they could touch it. This result is similar to other studies investigating the effects of agent embodiment on human-agent interaction. Lee et al. [1] suggested that a physically embodied agent may facilitate better social interaction with its users by providing more affordance for proper social interaction than a disembodied agent. In the same line, Kidd and Breazeal [5] indicated that the fact that people consider robots as "real entities" might facilitate face to face interaction.

V. CONCLUSION AND FUTURE WORK

In the current experimental setup, participants did not report any perceived added values of a robot and a virtual agent in comparison to a simple laptop PC in a specific interaction situation. This finding can be explained by the fact that in this experiment, the robot and the virtual agent lacked living characteristics and that the task required participants to focus on task performance rather on their interaction with systems. Future studies should address the issue of agent design and use different kinds of tasks to gain insight into the effects of agent embodiment on human-agent interaction.

Furthermore, in a future experiment, we should reduce or simplify interfaces of interaction systems because of divided attention difficulties in older adults. Finally, our findings are in line with those of previous studies, showing people prefer the physical embodiment of a robot rather than a projected bust of a virtual agent in an interaction situation.

ACKNOWLEDGMENT

This study has been supported by the French National Research Agency (ANR) through the TECSAN programme (ANR-09-TECS-012) and the CNSA (Caisse Nationale de Solidarité pour l'Autonomie).

REFERENCES

[1] K. M. Lee, Y. Jung, J. Kim, and S. R. Kim, "Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction," *International journal of human-computer studies,* vol. 64, pp. 962-973, 2006.

[2] K. Dautenhahn, "I could be You: The Phenomenological Dimension of Social Understanding," *Cybernetics and Systems: An International Journal,* vol. 28, pp. 417-453, 1997.

[3] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems,* vol. 42, pp. 143-166, 2003.

[4] J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric, "The role of physical embodiment in human-robot interaction," in *IEEE Proceedings of the International Workshop on Robot and Human Interactive Communication*, 2006, pp. 117-122.

[5] C. D. Kidd and C. Breazeal, "Effect of a robot on user perceptions," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004, pp. 3559-3564.

[6] J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric, "Embodiment and human-robot interaction: A task-based perspective," in *16th IEEE International Conference on Robot & Human Interactive Communication*, 2007, pp. 872-877.

[7] K. Shinozawa, F. Naya, J. Yamato, and K. Kogure, "Differences in effect of robot and screen agent recommendations on human decision-making," *International journal of human-computer studies,* vol. 62, pp. 267-279, 2005.

[8] T. Komatsu and Y. Abe, "Comparing an On-Screen Agent with a Robotic Agent in Non-Face-to-Face Interactions Intelligent Virtual Agents." vol. 5208, H. Prendinger*, et al.*, Eds., ed: Springer Berlin / Heidelberg, 2008, pp. 498-504.

[9] A. Pereira, C. Martinho, I. Leite, and A. Paiva, "iCat, the chess player: the influence of embodiment in the enjoyment of a game," in *The 7th international joint conference on Autonomous agents and multiagent systems*, 2008, pp. 1253-1256.

[10] D. Hasegawa, J. Cassell, and K. Araki, "The role of embodiment and perspective in direction-giving systems," in *Dialog with robots: AAAI Fall Symposium*, 2010.

[11] S. Pasquariello and C. Pelachaud, " "Greta : A Simple Facial Animation Engine"," in *6th Online World Conference on Soft Computing in Industrial Applications*, 2001.

[12] C. Jost, B. Le Pévédic, and D. Duhaut, "Creating Interaction Scenarios with a New Graphical User Interface," in *International Workshop on Intelligent Interfaces for Human Computer Interaction,* Palermo, Italy, 2012, pp. 531-538.

[13] P. Sweetser and P. Wyeth, "GameFlow: a model for evaluating player enjoyment in games," *Computers in Entertainment,* vol. 3, pp. 3-3, 2005.

[14] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, "Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model," *International Journal of Social Robotics,* vol. 2, pp. 361-375, 2010.

[15] M. Sarter and J. Turchi, "Age-and dementia-associated impairments in divided attention: psychological constructs, animal models, and underlying neuronal mechanisms," *Dementia and geriatric cognitive disorders,* vol. 13, pp. 46-58, 2002.

Augmenting Remote Trading Card Play with Virtual Characters used in Animation and Game Stories

- Towards Persuasive and Ambient Transmedia Storytelling -

Mizuki Sakamoto, Todorka Alexandrova, Tatsuo Nakajima
Department of Computer Science and Enginnering
Waseda University, JAPAN
{mizuki, toty, tatsuo}@dcl.cs.waseda.ac.jp

***Abstract*—Using well-known virtual characters is a promising approach to enhance information services, since such characters provoke people's empathetic feelings easily, and it is also easy for people to recall the leitmotif of the character's fictional stories. In Japan, recently, it has become a popular culture to use famous virtual characters of animations and games in various services, and this has even become a main business activity for some companies. In the real world, our daily life consists of various social activities, and virtual characters offer the possibility to enhance these activities. For example, our current social activities might be gamified by replacing unknown people with our favorite virtual characters or might be augmented by the characters' stories.**

In this paper, we present *Augmented Trading Card Game* that enhances remote trading card game play with virtual characters used in the fictional stories of popular animations and games. We show our observations about the way players use the system, realizing the game, and what their feelings and impressions about the game are. We believe the obtained results would be useful to consider how to use empathetic virtual characters and the fictional story that the characters are used in, in the real world activities for future information services. We also discuss how our approach can be extended to design a new type of transmedia storytelling by considering *Augmented Trading Card Game* as one form of transmedia storytelling.

Keywords - Empathetic virtual characters; Game design; Augmented reality; Trading card game; Ideological metaphor; Animation and game stories; Physical tangibility; Transmedia storytelling

I. INTRODUCTION

Virtual characters are widely used in our daily life, recently. For example, famous Disney characters like *Mickey Mouse* and *Donald Duck* provoke our empathetic feelings easily anytime and anywhere, and *Kitty* and *Pokémon* are now found all over the world [1]. In animations and games, each virtual character has its own personality and story, which can be used as a medium to convey special information and messages to people. If people are familiar with the fictional story of an animation or a game, then the story's characters are able to recall the leitmotif of the story easily without much additional information but by just performing some action/interaction with the story's character. Specifically, many Japanese animation and game stories contain serious ideological messages that are important to make our daily life more desirable. We believe that this observation is very important when considering how to use virtual characters in various future information services in the real world. Recently, our daily life is becoming more and more complex, and we need to process a lot of information everyday. In our modern society, there are many real social problems that need to be solved urgently. However, it is hard to convey important ideological messages to people without presenting a large amount of additional information to them. For example, education is a traditional heavy-weighted method to teach the importance of ideological messages without long learning time.

In the contemporary Japanese society, several posters for public service announcements have adopted the idea to use virtual characters from recent popular animation stories. For example, *K-ON!* has been used for promoting a national survey in Japan [24]. In the story of *K-ON!*, high school girls try to realize their dreams with cooperative efforts among them. This becomes a persuasive message conveying the idea that everyone's participation is important for the national survey. Also, NFGD that promotes guide dogs' training has created two posters using popular characters from *Puella Magi Madoka Magica* [26]. The girl that has been used in one of the posters is rebellious, but very considerate to her friends. The girl used in the other poster is very close to her friend and keeps thinking and caring of her friend even when they are far from each other. Many young girls admire these two girls, nowadays. If people want to imitate these magic girls, then the posters contain the implicit, strong persuasive message that becoming a puppy walker is one of the ways to become like them. These examples show the effectiveness of using virtual characters that have their background stories to attract people. Moreover, they are a good evidence that virtual characters could be used to convey ideological messages that might play significant role in changing people's current attitude. In Japan, the majority of young people have been enjoying animation and game stories for a long time and they know the popular animation and game characters and their stories very well. This we believe is a good prerequisite for using virtual characters to enhance emotional feelings and successfully convey ideological messages through the characters' stories.

This paper presents *Augmented Trading Card Game* (*Augmented TCG*) for playing the *Yu-Gi-Oh! Trading Card*

 ISBN: 978-1-61208-250-9

Game (*Yu-Gi-Oh! TCG*) between two players who are located in different places. The system, realizing the game, supports the remote trading card game play with virtual characters. We consider two possibilities to use virtual characters in the game. The first one is to show a virtual character representing the opponent player, where the movement of this character is synchronized with the current real movement of the opponent player. We believe that this movement synchronization gives the character more reality, which allows us to easily remind the players the leitmotif of the characters' stories containing some ideological messages. The second idea is to include another virtual character, which is actually the virtual character of the character depicted on one of the player's cards. This character would encourage the player to win during the game. Some *Yu-Gi-Oh! TCG* players are very keen on collecting cards and they feel strong empathy with their cards. Moreover, they even feel the characters illustrated on their cards as their close friends. Therefore, the communication with the characters during the play would have a very strong influence on the players.

In *Augmented TCG*, we use popular virtual characters that have been used in Japanese animations and games. Specifically, the *Yu-Gi-Oh! TCG* has been originally introduced in the *Yu-Gi-Oh!* comic and animation. One of the reasons why *Yu-Gi-Oh! TCG* is popular in Japan is the fact that almost all young people have first enjoyed the comic and the animation story and then learnt how to play the game from that story. The story also teaches some important ideological concepts such as the importance of justice, friendship, bravery, positivity, and thoughtfulness. That is why we believe that the characters of the *Yu-Gi-Oh! animation* story can be used to enhance the playing style of the game through the stories they carry and recall. In this research we are interested in investigating the impact of the presence and behavior of the 3D virtual characters on the players' emotions and feelings, and the play style of the game. For that purpose we conducted a user study in which we observed the participants' attitude during the play, and interviewed them after that. Furthermore, we discuss that *Augmented TCG* is considered as one form of transmedia storytelling. Transmedia storytelling is the technique of telling a single story or story experience across multiple platforms and formats using current digital technologies [4]. Our approach enhances an original fictional story in corporation with a real world game or social information service to create a new story that a player of the game or a user of the social information service participates in. We also extract the basic concepts from *Augmented TCG*, and present a design framework for designing a new form of transmedia storytelling, and as a next step we would like to investigate the effect of using virtual characters from animation and game stories in our daily activities.

The remaining sections are structured as follows. Section II shows an overview of *Yu-Gi-Oh! TCG*. Section III presents some issues for designing remote trading card play. In Section IV, we present *Augmented TCG* and describe its experiments. Section V discusses that *Augmented TCG* is considered as a new form of transmedia storytelling and our design framework for designing a new form of transmedia storytelling. In Section VI, we show some related work of *Augmented TCG*. Finally, Section VII concludes the paper.

II. YU-GI-OH! TRADING CARD GAME

The *Yu-Gi-Oh! TCG* [29] is a trading card game based on the Duel Monsters game that is portrayed in the popular *Yu-Gi-Oh! comic*. *Yu-Gi-Oh! cards* are categorized into three types: *Monster*, *Spell* and *Trap* cards. A *Yu-Gi-Oh! TCG*'s player structures his/her own original deck by selecting his/her favorite cards from the several thousands *Yu-Gi-Oh! cards*, currently released. This leads to each user having his/her own unique and original deck that reflects his/her own personality and taste.

Figure 1. Playing Trading Card Game.

Yu-Gi-Oh! TCG is a turn-based game, which is played in a one-to- one or two-to-two manner as shown in Fig. 1. We call the battle with *Yu-Gi-Oh!* cards a *duel*. Each player starts the game with a certain number of points called *life points* and performs the duel by summoning his/her monsters, battling against the opponent with his/her monsters or using spells and traps. Depending on the action taken and the outcome of it, the *life points* decrease or increase. If the *life points* of one of the players become zero or he/she cannot draw cards from his/her own deck, then that player loses the duel and the game ends.

The *Yu-Gi-Oh! TCG* involves various sources of enjoyment, besides just playing the game, such as completing collections of cards, structuring decks, communicating with the opponent players, trading, battling, as well as establishing different links to *Yu-Gi-Oh! TV animations* and *Yu-Gi-Oh! comics*. Moreover, although computerization has advanced a lot currently, the traditional version of *Yu-Gi-Oh! TCG* that does not use a computer but paper cards is still very popular among players.

III. REMOTE TRADING CARD PLAY

In Japan, as already described, trading card games like *Yu-Gi-Oh! TCG* are very popular among children, but most of them stop playing the games when they become adults since they have no time to meet and play with other people any more. However, because of the complicated rules and interesting strategies, trading card games are still very attractive and popular among adults. That is why playing a trading card game with a remote opponent player has become a typical play style in the busy modern society. In many other games like *Go*, *Chess, Poker* and *Mah-jong*, the remote play style becomes more and more popular as well. In this section we describe some problems and limitations related to the existing remote trading card play.

There are already several systems that support remote trading card play. *Yu-Gi-Oh! TCG* on *Nintendo DS* uses *Wi-Fi* to connect remote players. However, in this version, the trading cards are digitally represented and thus, players do not have the sense of the physical tangibility of the cards, which decreases significantly the pleasure and enjoyment of the game for some of them [16]. Specifically, the physical tangibility of the cards is essential to make players feel empathy with their cards. *Yu-Gi-Oh! Online* also offers a similar play style on personal computers, and suffers from the same flaw. Therefore in order to improve the described limitation of the existing systems, in *Augmented TCG*, we allow players to use their own, already collected physical paper cards.

As described in [16], we have found that some players would feel uncomfortable communicating directly with a real opponent player when the opponent player is a stranger for them. However, using just an avatar to represent the opponent player, as in the existing online trading card games, does not offer enough reality for the opponent player. We believe that using virtual characters from animation and game stories to represent the opponent players, whose behavior is synchronized with the behavior of the opponent player, would offer more reality than the avatars used in the current online trading card games. Also, trading card games in Japan are very closely related to comic, animation and game stories, in which the trading card games are part of the stories. That is why players usually relate to these stories and their characters while playing the game, which makes the game even more enjoyable for them.

In the current online trading card games, some players tend to disconnect from the game if they are going to lose. Specifically, it is a typical behavior of players whose only goal is winning the game. Such players feel pleasure and satisfaction only when winning the game and do not just enjoy the gaming process. This "unfair play" becomes one of the reasons for many trading card game players to dislike playing the online versions of the trading card games [16]. However, even unfair players know the comic, animation and game stories well. As described before, the characters in these stories convey ideological messages like the importance of friendship and the pleasure of honest and fair play. Involving these characters in the game would remind the players these ideological messages and encourage their desirable fair play style.

Many Japanese young adults admire animation and game stories and this strong worship has started and developed since they were children. If some of the hurdles to use remote trading card play vanish by solving the problems described in this section, many people will continue playing the game even when they become adults. If young adults keep their passion on the Japanese modern culture, this will become a power to successfully change our current undesirable social situations.

IV. AUGMENTED TRADING CARD GAME

A. Overview of Augmented Trading Card Game and its Experiment

Augmented TCG enhances the remote trading card play performed by two persons. The basic design approach is similar to the one of the augmented reality games introduced in [19], which integrates physical items and virtual items. As shown in Fig. 2, the two players are located in different places. Each player's cards, in his/her duel field on the table in front of him/her, are captured by a camera and projected on the opponent player's table.

Also, each player is represented by the 3D model of a virtual character used in popular animations and games, and this character is shown to the opponent player. In the current implementation, *MikuMikuDance* [25] is used to show the 3D models of the virtual characters. *MikuMikuDance* is free software for creating 3D movies by using virtual characters. The virtual character is controlled using *MS Kinect* and its movement is synchronized with the movement of the opponent player. In the current *Augmented TCG*, a player can choose one of three virtual characters that are *Yugi* and *Kaiba* from the *Yu-Gi-Oh!* animation story, and *Link* from The *Legend of Zelda* [27]. In the *Yu-Gi-Oh!* animation story, *Yugi* is always surrounded by many friends and his winning success is a result of his strong bonds with his friends who love the trading card game. *Kaiba* is a lonely hero and he always seeks the strength in the game, but he does not accept other people's help even if he is in a critical situation. However, in the story he also finally understands the importance of friendship. Most young boys want to follow either of these two characters because of their typical, very attractive and ideal personalities. The reason to choose *Link* as the third character in our experiment is that we would like to investigate how a popular character from another unrelated to TCG story affects the attitude of a player.

Furthermore, while playing the game, another virtual character, that has been depicted on one of the player's cards in advance, appears on a small display near the player once that card is drawn out of the deck, and supports and encourages him/her to win the game until the end of the game.

Moreover, the two players can communicate with each other via Skype if desired, and thus it is possible for them to introduce each other directly instead of using virtual characters. This option will allow us to compare the case of direct communication and the case of communication through virtual characters between the players.

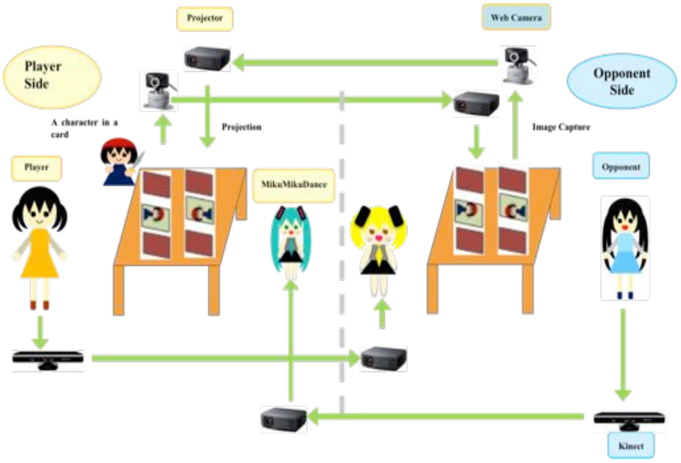


Figure 2. Augmented Trading Card Game System.

For the experiments, described in the paper, the rules of the game have been simplified for making the duels shorter and special predefined decks of cards have been used. The decks have been prepared by one of the authors who has more than 10 years experience in *Yu-Gi-Oh! TCG* and has more than few thousands *Yu-Gi-Oh! trading cards*. She is very familiar with the TCG animation story, and knows well how each character structures the deck and uses the cards in the animation. Therefore, for each possible virtual character to be chosen by a participant to represent him/her in the game, a suitable deck consistent with the animation story situation has been prepared. In the current version of *Augmented TCG*, the virtual character's behavior does not reflect the real behavior of the player precisely, but the behavior is exaggerated and overreacted according to the current play situation. One of the reasons for this is the fact that our system's functionality implementation is not currently completely finished. That is why in the current system another person exaggerates and overreacts the behavior of the player and the movement of the virtual character, representing the player is synchronized with that behavior using MS Kinect. This is also done in order to make the movement and behavior of the virtual character closer to the actual character's behavior in the animation.

Fig. 3 shows the current prototype configuration for a participant. On a large display, a virtual character, which movement is synchronized with the movement of the person who imitates the opponent player, is shown. A camera is setup behind the small display near the participant, and captures the image of the cards. The opponent player's cards are projected on the table by a projector. A small display shows the other virtual character that is depicted on one of the player's cards, which in this case is one of the most powerful cards in the participant's deck.

We recruited five participants for our experiments, and they all performed the duels in the experiments against one of the authors of the paper, who has deep knowledge about the TCG and could lead and control the experiment so that all participants play the game under the same conditions. Most of the participants had more than three years experience in *Yu-Gi-Oh! TCG* and they knew the characters in the animation stories very well. Before the experiments players could not talk to each other and none of them knew about *Augmented TCG*. Also, they were told how the rules were simplified right before the experiment. During the experiments, we observed each participant's play and conducted interviews with him/her after the play based on the contextual inquiry method [3]. The experiments are described in details in the following two subsections.

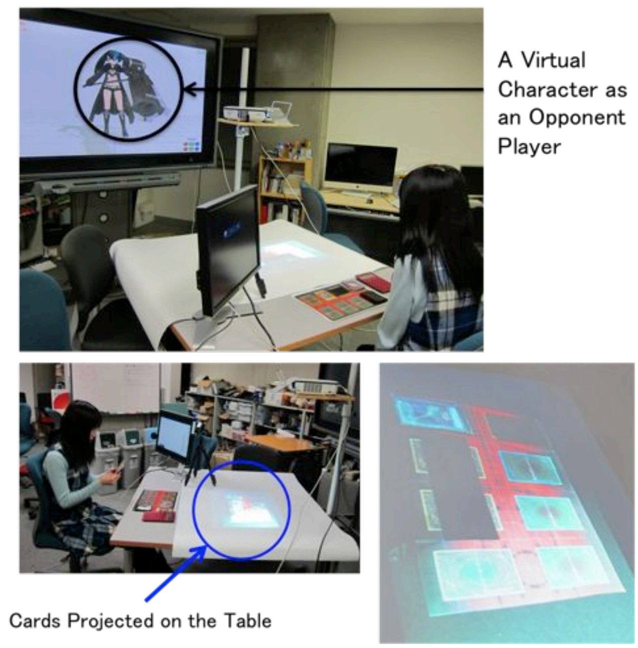


Figure 3. Current Prototype Configuration.

B. *Playing against a Virtual Character*

We did two experiments for playing *Augmented TCG* against a virtual character representing the opponent player.

In the first experiment, participants could choose either *Yugi* or *Kaiba* for their own character according to their preferences. After the play, we interviewed the participants about their impressions about the virtual character representing their opponent. One of them said: *"I could feel I am playing against Yugi, but Yugi used in the experiment does not offer enough reality"*. Especially, the movement of the character was sometimes not like the real movement of *Yugi as in the animation story*. He also said: *"I will definitely enjoy more the game against Yugi, and would like to win the game if the movement is more realistic"*. Another participant said: *"The face expression of the character is* *poor and it is a very important issue while playing a game against a real person"*. Also, one of the players told us: *"The voice should be the same as the actor's voice of the character in the animation story"*. Moreover, if the opponent player was a female, some participants felt strange since both *Yugi* and *Kaiba* were male characters.

In the animation story, players usually play *Yu-Gi-Oh! TCG* standing, that is why we chose that the characters were always standing during the play, but in the real situation, players usually play sitting. Some participants felt the unreality on the behavior of the characters, but if the characters just sat down all the time, the participants also felt the inconsistency with the *Yugi* and *Kaiba*'s personality.

Figure 4. Encouraging a Player by a Virtual Character.

In the second experiment, *Link* from *The Legend of Zelda* was used as a character representing the opponent player. The results in this case were completely different depending on whether the participants liked this character or not. If the participants were not interested in *Link*, they usually did not care about the presence of *Link*, but if *Link* was their favorite character, then they found playing the game against *Link* more enjoyable. One of the male participants also told us: *"If the character is a pretty girl, I may be more excited to play the game"*. Also, a female participant told us: *"I feel that Link is my boy friend, so playing against him increases my pleasure and positivity"*.

C. Encouraging a Player by a Virtual Character

In this experiment, a small display on the table showed the virtual character depicted in advance on one of the cards used by the participant, as shown in Fig. 4. We have selected *Dead Master* from *Black★Rock Shooter* [20] as a character to be depicted on the card because we feel that the character does not contradict with or violate the atmosphere of *Yu-Gi-Oh!*. *Black★Rock Shooter* has two worlds. *Dead Master* is an enemy of *Black★Rock Shooter* in another dimension world, but in the daily world, they are very close friends. This becomes a persuasive message conveying the meaning that players need to keep and develop their friendship even if they fight seriously in a game.

As already described, in this experiment a special deck was structured in advance for each participant depending on the character he/she chose to play with. Then, in the duel, the participant always drew out of the deck the card depicting *Dead Master* at the beginning of the game. Once that card had been drawn out, a small display next to the player kept showing *Dead Master* until the end of the duel. *Dead Master* supported and encouraged the player during the game by using encouraging body gestures and its movement was controlled by a person who operated *MS Kinect*.

After the experiment one of the participants said: *"It is desirable that the card depicting Dead Master does not lose from the attack of the opponent player"*. However, another player who was not interested in the character told us: *"It is more enjoyable if the participant's favorite character encourages him"*. One of the other participants said: *"I feel that the character does not encourage me enough using only gestures. It is better that the character talks or advises me"*. He also told us: *"It is desirable that the character behaves like a cheerleader"*. *Dead Master* is a serious character, so if that character just behaves as a cheerleader, some players who know the animation story of *Dead Master* may feel the unreality due to the loss of the consistency with the story. Also, another participant told us: *"The encouragement should be like the one in the animation story"*. Most participants said: *"The presence of the character increases the pleasure, but it is hard to consider winning the game just from that encouragement"*. The participants' comments showed that they were aware that exactly the character depicted on one of their cards appeared on the small display without them being informed in advance about this feature of the system.

D. Discussions

The experiments described in the previous subsections have shown that using virtual characters affects positively the play style and enjoyment of the game. However, they also show that reality is an important issue in order to successfully include virtual characters and as a byproduct to recall the leitmotif of their stories in the play. In this subsection we give a discussion on some observations obtained from the *Augmented TCG* experiments.

As already mentioned, most of the participants, who have watched the *Yu-Gi-Oh! animation* story, feel that adding popular virtual characters from the animation to the game makes it more enjoyable and exciting. The desire to follow these characters becomes also an incentive for the users to change their behavior. A negative feedback may be used to achieve moral play, but changing a user's attitude in general is not easy [13]. The most important insight is that there is a possibility to use a virtual character as a metaphor that recalls the story of the character in the player's mind while gaming, and in this way the story may convey the leitmotif containing ideological concepts like the importance of friendship, honesty, thoughtfulness and so on. This approach would have the power to change the player's attitude.

During the experiments most of the participants enjoyed playing a game against a virtual character, and being encouraged by a virtual character. However, as already mentioned from the experiments and the interviews with the participants we have found that the reality of the characters, such as facial expressions, movement and behavior, is essential to increase the enjoyment of the game. Reality is an important criterion to evaluate a design [15]. In our case, the virtual character's behavior should be consistent with the character's behavior in the animation story and this is important to feel the reality in the game. For example, cards that are not used by the character in the animation should not be used in the game as well, and the movement of the virtual character should be consistent with its typical movement in the animation. The number of cards owned by the character should be consistent with the real game situation as well. A gap between the reality and the virtuality may cause the *uncanny valley* problem [10]. In the second experiment described in Section IV.C, the participants are not very familiar with the chosen virtual character but it may cause a feeling of incongruity in the players while playing *Yu-Gi-Oh! TCG* if they are aware that the character appears in another animation story that is unrelated to TCG.

The story behind the characters gives some influence on players' attitude as well. Since the influence of the story behind the character is especially interesting in our research, we discuss the issue in more details below.

All participants in the experiments could easily recall the *Yu-Gi-Oh!* animation story during their plays. The *Yu-Gi-Oh!* animation story contains some ideological ideas, such as the importance of the competition and cooperation among friends for their growth, but the current story may not be enough powerful to remind its ideological idea to players during their play because the growth of the main hero due to the competition and cooperation among friends is rather implicit in the story. However, the story is able to increase the positive attitude while playing the game, and to increase the self-efficacy to win the game. We believe that this factor is important to enhance our approach. In our daily life, people may not feel self-efficacy to change undesirable attitude and behavior because our daily life becomes more and more complicated, and we do not have enough time to consider the importance of the desirable behavior. We also need to consider how the representation form of the stories affects the conveying of ideological idea in the near future. In Japan, the same story is represented in different forms such as animation, manga, game, and novel.

In [13], several persuasive services have been reported to change people's undesirable behavior. Although these services successfully change people's undesirable behavior temporally, it is not easy to change their attitude to maintain their desirable behavior for a long time. Using virtual characters and their stories offers a new possibility to enhance the previous approaches by enhancing the human positivity.

Most people, especially Japanese, like fictional stories that bring them positive feelings with the fact that people feel able to do whatever they want in the fictional stories by using hidden magic abilities in the fictional world. If a game becomes more pervasive into our daily life, the boundary between the real world and the fictional world would become more blurred. Therefore, fictional stories can be used to enhance our daily activities. Using the stories increases the human positivity in the real world, and enhances people's self-efficacy to do what they want because through the games they can have the illusion of having special extraordinary abilities that exist in the fictional world. This is really true for all Japanese young people who have grown up with fictional stories like game, animation and fictional stories that use special effects. Positivity is an important fact to increase people's self-efficacy as proved in the positive psychology research [8]. The positivity is very useful to make our life more meaningful [11], and the positivity is essential for the success in our life. Our approach is one possible powerful way to enhance people's positivity.

Specificlly, many Japanese animation and game stories emphasize on the importance of positivity. Thus, the characters in the stories can be used as metaphors to increase human's positivity while playing a game. It will become one of the most important roles of a game to teach solutions to typical serious problems in life.

Another worth mentioning observation is that in the current experiment players could not see their own virtual character during the play, but only the opponent player's character. We consider that adding this feature to the game would increase the pleasure and the reality of the play since users will be able to control naturally the behavior of their own virtual characters.

V. TOWARDS PERSUASIVE AND AMBIENT TRANSMEDIA STORYTELLING

A. *Augmented TCG as Transmedia Storytelling*

Enhancing games played in the real world like TCG with fictional stories is a promising direction to design a new form of transmedia. In this section, we discuss some design implications of *Augmented TCG* that will be considered as one form of future transmedia storytelling [4].

In the current *Augmented TCG*, its animation and game story is explicitly not shown during the play. A player needs to recall the story during his/her play. More tight integration of the game play, and the animation and game story offers a new possibility to design transmedia. The movement of a virtual character from the virtual world to the real world offers a tight integration between the fictional story and the

TCG game play. Transversal interfaces [2] offer a way to move between the worlds seamlessly. The approach offers a stronger association between a fictional story in the virtual world and the real world through a virtual character than the current approaches, and the boundary between the two worlds becomes more blurred.

We believe that a participant would be more excited to play the game if a character drawn on a card from his/her favorite deck is shown to encourage him/her. Cards are considered as one piece of transmedia to construct a fictional story. However, the preferences for that character are different according to the player's gender. A female player usually likes a card depicting a pretty girl. In this case, encouraging cheerfully the player with gestures by that character would be natural and meaningful. On the other hand, a male player usually likes a powerful card that may depict a strong monster. In such a case, the encouragement by the character should be more powerful and adding special effects to show the superior ability of the character would be more suitable and more effective to motivate the player. Thus, if the character is one of the player's favorite characters, then the encouragement would be a powerful tool to increase the player's motivation and excitement of the game.

When playing with a virtual character from animation and game story, the player also tries to mimic the character's behavior in the animation story. This can be a very useful and successful approach to teach players how to improve their gaming skills. If players follow the skillful character's way of playing in the story, then they can learn new skills and techniques from that character's experience in the animation. Of course, a skillful friend is a good coach for improving a player's skills, but if there is no good coach available around the player, then he/she needs to learn by himself/herself and doing it following the experience of the character would be a promising and exciting approach to exploit future transmedia storytelling.

In our experiments, we could not find the rigorous evidence that the stories of the virtual characters could always strongly affect the attitude of the players. One of the reasons is that in our current research we focus on a game. For most people, the purpose of a game is just for fun. Of course, the duel against *Yugi* and *Kaiba* makes players play the game more seriously, but it is hard to make players braver when *Link* is shown as an opponent player. *Link* is a character in an RPG game, and a male player considers that the character is just like his avatar in the game, so his story does not have strong ideological messages in the game. Also, the presence of *Dead Master* does not have a strong impact on a player, since the character itself is very popular, but its story is not so well known yet. This means that well known stories that contain strong ideological messages and characters that have powerful and distinctive personalities are important to make virtual characters be used as metaphors. We also consider that the music used in the popular stories could also become a metaphor for the stories because in Japanese animations, their music sometimes becomes more popular than their characters. We believe that designing metaphors that use the popular stories in animations and games is a promising future direction to convey complex ideological messages to people without presenting a large amount of information to them.

One of the problems in using virtual characters is their copyright. There are many free 3D models for *MikuMikuDance*, but some of them are deleted on the Web due to the copyright issues. However, freely available models offer new possibilities to enhance games because the models can be easily customized. In Japan, it is a popular culture to create new characters and stories from existing ones. Using a customizable virtual character in *Augmented TCG* may create a new playing style of TCG, and the new stories of the character can be used to enhance its role as a metaphor.

As already described, virtual characters used in animations and game stories are widely used in multiple media channels. In *Pokémon,* a synergy among games, movies, and TV programs is used to make the *Pokémon* story more popular, and make the story pervasive in its fan's daily life. Also, in the *Yu-Gi-Oh! animation* story, the animation story teaches its game players how to play the TCG game and why the game is attractive while they are watching the animation story. Using multiple channels to communicate messages among people is a very effective way to convey the messages among people because each channel can convey the message in a special partial way. This is also a typical approach in the current advertisement because only one medium cannot deliver the advertisement to a large audience of people.

B. Value-Based Design Framework

Fig. 5 shows a framework for designing persuasive and ambient transmedia storytelling. In [15, 17], we extracted the values that are useful to design attractive services and products from our previous case studies. The classification of the values is extended for designing persuasive and ambient trasmedia. Our framework consists of two components. The first component is a fictional story, and the second component is a pervasive game or social information service. It offers its own goal for a player. For example, in *Augmented TCG*, the goal is to win a duel, to collect TCG cards, or to enjoy a duel with friends [16]. The four values depicted in Fig. 5, informative value, empathetic value, economic value and aesthetic value are used to increase both extrinsic and intrinsic motivation to achieve the goal of the pervasive game. The fictional story also offers the four values that are the same as in the pervasive game, but it also offers two more values, a positive value and ideological value. The positive value in the fictional story offers people a feeling that increases their positivity. This finally increases the self-efficacy to aim the achievement of the goal identified in the ideological value. Many Japanese game and animation stories depict the

 ISBN: 978-1-61208-250-9

heroes' or heroines' final success. In the stories, they never give up until achieving their goals, so it increases human positivity. The ideological value offers the important insights to consider people's life. For example, Japanese games and animation stories like to teach us various ideological concepts such as justice, human growth and development, co-existence with human and nature, love, and friendship. Incorporating the ideological value in the stories makes it possible to educate people to understand the importance to sympathize the stories' goals, and motivate them to try to achieve the goal.

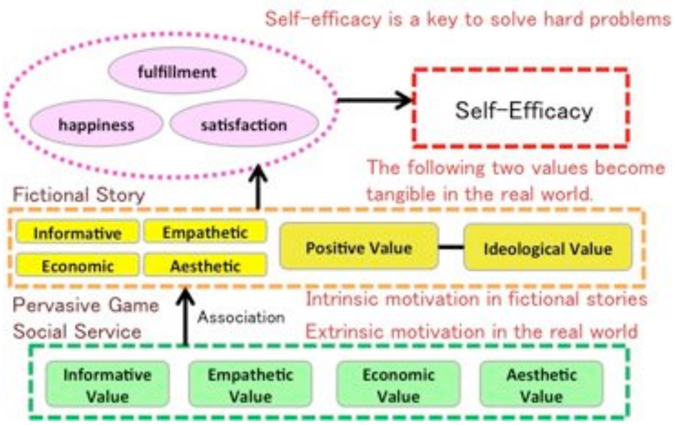


Figure 5. Value-Based Design Framework for Designing Persuasive and Ambient Transmedia Storytelling.

In our framework, we are considering two types of storytelling. In the first type, the ideological value in the fictional story becomes a main issue in the storytelling. For example, *Augmented TCG* adopted this type's storytelling. The pervasive TCG game offers four values to make a user achieve the goal of the game by increasing his/her intrinsic and extrinsic motivation. As describe above, a fictional story offers the positive value and the ideological value. While a player is playing a game, the game offers some metaphors of the fictional story to remind the story to the player. In *Augmented TCG*, using *Yugi* and *Kaiba* as an opponent player reminds the player the *Yu-Gi-Oh* fictional story. The virtual characters create associations between the four values offered by the pervasive TCG game and the fictional story. The association makes a player feel the fictional story as a more realistic story in the real world. Thus, he/she considers that the positive value and the ideological value in the fictional story also exist in our real world. The ideological value that is tangible in the real world helps a player to recall the importance of the ideological concept presented in the fictional story, and motivates him/her to achieve the goal identified in the story. For example, in *Augmented TCG,* a player recalls the importance of justice and friendship, and then he/she considers playing the TCG game more fairly, and growing as a noble person. Also, the tangible positive value in the real world makes a player increase his/her self-efficacy to have enough confidence for achieving the goal of understanding the importance of the ideological concept. This approach offers possibilities to solve complicated social problems and to implement various public policies [30].

In the second type of storytelling, the goal defined in the pervasive game or social information service is more important. The most important aspect of this type is to define the goal that is not related to the ideological values incorporated in the fictional story. For example, we can establish the goal of the pervasive game or social information service to be to encourage people to visit a certain restaurant more frequently. In this case, the four values offered by the pervasive game or social information service become extrinsic motivation to achieve the goal. We can use some metaphors of the fictional story, such as a virtual character from it, as an association to remind the story to the player while playing the game. Then, the positive value in the fictional story would make a player feel self-efficacy to achieve the goal in the pervasive game or social information service. Finally, the ideological value makes people feel that the goal is more desirable and increase the priority to achieve the goal in their daily life.

VI. RELATED WORK

There are several other systems that support remote TCG play. *Duel Accelerator* [22] is an online-based *Yu-Gi-Oh! TCG* where each player chooses his/her avatar and virtual trading cards with special effects shown on them during the duel, are used. The special effects become emotional stimulus for the player and thus the pleasure of the play is increased. However, virtual cards loose the sense of the physical tangibility of the cards, and it is hard to motivate a player to enjoy collecting cards. *The Eye of Judgment* [23] uses augmented reality technologies to show special effects on the real trading cards. It allows players to use real physical cards and special effects can be shown on them. In *Augmented TCG*, it is easy to add special effects to the cards projected on the table in front of a player. Also, the *Skype duel* uses Skype to show each player's cards on the opposite player's display and the voice communication between the two players is possible as well. *Augmented TCG* uses 3D virtual characters that are shown to the players and the characters move according to the opponent player's current behavior.

In *CyberOne* [21], which is a new TCG, each paper card has a sequence number. When the number is input in the system, the corresponding virtual card appears in the online TCG. The player can exploit the tangibility of the cards, but he/she can also enjoy additional special effects, which is an example of the advantages of the virtual cards. Once the set of cards in the player's deck are input in the system, the duel is performed automatically without the player doing anything and the final result of the duel is returned to the player, showing the strength of his/her deck selection. However, the enjoyment of constructing an original deck still remains for the player. Moreover, the two players do not need to play the game at the same time because they just need to construct their decks.

Augmented reality techniques may be used to enhance existing games. For example, in [19], several augmented reality games, that enhance traditional games to integrate physical items and virtual items, are reported. *Augmented Go* is an example of a promising approach to maintain the advantages of the physicality, but to add virtuality to the Go board game. Also, in [6], a card game is enhanced by embedding RFIDs in the cards to identify which card is currently used, and offer appropriate information to players.

Designing a user's experience [7] is related to the value based design framework. In the current approach, we extracted and analyzed values from a newly designed artifact, but it may be desirable to investigate how the values are emerged by analyzing a user's activities. Since designing values strongly depends on a user's experience, in the next step, it is an important research question to explore how values affect the user's experience, and how the values designed in the artifacts interact with the values emerged in the user's experience.

Popular Japanese role playing games such as *the Legend of Zelda* and *Kid Icarus* offer rich fictional stories incorporating the positive and ideological value. However, the games cannot make the values tangible in the real world because the games do not offer associations between the real world and the virtual world.

The Alternative Reality Game (ARG) [12, 18] is a promising approach to convey messages to people using multiple channels. Fictional stories are embedded into a pervasive game that can use multiple channels. The channels offered in the game are used to exploit the game's fictional story. For example, in Perplex City [28], trading cards are used to introduce its characters and story. Web sites, emails, phone calls, and SMS messages are cooperatively used to solve riddles in the mystery story. Our approach is also a useful way to design the transmedia story telling [4]. Especially, the form to represent a story affects how an ideological message is conveyed to a user. The transmedia storytelling divides a story across multiple media so that it is a possibility to choose the most suitable form to attract a user and to convey a message to a user.

"Seichi Junrei" is a typical geek culture in Japan, especially related to Japanese animation, manga and game, in which people tend to visit famous locations from animation, manga and game. *"Seichi"* means "*Holy Land*", *"Junrei"* means "*Pilgrimage*". Anime fans arrive at that location, and take pictures with the same screen/angle of the animation, and upload them to their blogs. The most important aspect of *"Seichi Junrei"* is to bring something from the fictional story to the real world. The fans create new stories with these pictures and the virtual characters appearing in the fictional stories, and co-construct the stories to share them within their communities. This is a very interesting phenomenon to harmonize the real world and the fictional world. We believe that interactive pervasive games or social information services based on fictional stories are very promising tools to increase the reality of the fictional world, and the tools enhance the *"Seichi Junrei"* phenomena by realizing more tight integration between the fictional world and the real world. The experiences described in the paper will offer useful insights to design tools that will realize new types of transmedia storytelling.

There are several psychological theories that are useful to design attractive persuasive and ambient transmedia storytelling that influences a user's behavior and attitude [14]. Burrhus F. Skinner's behavior science makes it possible to change people's behavior and attitude unconsciously. *Operant conditioning* is a form of learning during which an individual modifies the occurrence and form of its own behavior due to the reinforcement of consequences of the behavior. Positive reinforcements and negative reinforcements construct a feedback loop to control human behavior and attitude systematically. Most games use the approach to fascinate people to play games. The *elaboration likelihood model* explains the importance for the dual routes to persuasion. The central route offers heavily cognitive information to change the user's attitude, and the peripheral route enables the user to change his/her attitude through emotionally influential information. The *transtheoretical model* proposes five stages as a process involving the progress to change the user's undesirable behavior. In earlier stages, the user prefers emotional reinforcement not to give up his current efforts. On the other hand, for the user who is in a near final stage, enough information for making a better decision through rational thinking is more suitable. The *feeling as information theory* is useful to consider how information technologies evoke the user's emotion. The theory indicates that it is difficult to think rationally during a positive feeling. On the other hand, the user tends to think rationally when she feels to be in a negative situation. The results indicate that positive stimuli are effective in early stages, but in latter stages, negative stimuli are desirable under the *transtheoretical model*. *A theory of unconscious thought* shows that heuristic thinking is not always bad. Giving more information may lead a wrong decision. When we become conscious, our heuristics is biased according to the current frame. However, as described in the theory, when our thought is unconscious, heuristic thinking may lead to a right decision.

VII. CONCLUSION AND FUTURE DIRECTIONS

This paper presents some observations on the usage and the design of *Augmented TCG* that enhances remote trading card game play with virtual characters used in animation and game stories. We have shown some experiments of *Augmented TCG*, and presented its design implications.

Currently, we are considering the following three future research directions. The first one is related to the physical tangibility. In *Augmented TCG*, the physicality of cards is one of the most important design decision, since virtual cards lead to the loss of the reality in *Yu-Gi-Oh!* TCG. Also, physical tangibility is important to offer more pleasure to a player [5]. However, virtual cards have some advantages over real paper cards such as the possibility to easily add special effects to the card. In the current *Augmented TCG*, a player needs to manually teach the system which cards he/she would like to use. If the system could automatically detect the player's cards, the usability of the system would be dramatically improved. The automatic detection of cards can be easily realized by inserting RFID in each card similarly to the approach described in [6]. In our current system it is sometimes hard to see clearly the opponent player's card and in such case a player needs to explain which card he/she is using and what effect the card has. If the detailed information about the card is automatically shown on the duel field, players do not need such extra communication by using a voice communication system or a chat system. Also in *Augmented TCG*, it is easy to detect that a player mistakes the usage of a card, which is very helpful since sometimes it is hard to understand the complex rules of a game even for a semi-expert player. The tangible interaction approach is very promising for improving the disadvantages of the virtuality [9], and makes it possible to combine successfully the advantages of virtual cards with the advantages of the real paper cards.

REFERENCES

[1] Anne Allison, "*Millennial Monsters: Japanese Toys and the Global Imagination*", University of California Press, 2006.

[2] Elisabeth Andre, Klaus Dorfmuller-Ulhaas and Matthias Rehm, "*Engaging in a Conversation with Synthetic Characters Along the Virtual Continuum*", In Proceedings of 5th International Conference on Smart Graphics, pp.1-12, 2005.

[3] Hugh Beyer and Karen Holtzblatt, "Contextual Design", Morgan Kaufmann, 1999.

[4] Christy Dena, *"Transmedia Practice: Theorising the Practice of Expressing a Fictional World across Distinct Media and Environments"*, Dissertation Thesis, University Sydney, 2009.

[5] Hiroshi Ishii, and Brygg Ullmer, "*Tangible bits: towards seamless interfaces between people, bits and atoms*", In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.234-241, 1997.

[6] Christian Floerkemeier and Friedemann Mattern,, "*Smart Playing Cards – Enhancing the Gaming Experience with RFID*", In Proceeding of the 3rd Int. Workshop on Pervasive Gaming Applications, , pp.79-88, 2006.

[7] Jodi Forlizzi and Katja Battarbee, "*Understanding Experience in Interactive Systems*", In Proceedings of the 5th International Conference on Designing Interactive Systems, pp.261-268, 2004.

[8] Barbara L. Fredrikson, ``*Positivity: Top-Notch Research Reveals the 3 to 1 Ratio That Will Change Your Life*", Three Rivers Press, 2009.

[9] Patrick W. Jordan, "*Designing pleasurable products: An introduction to the new human factors*", Routledge, 2002.

[10] Masahiro Mori, "*On the Uncanny Valley*", In Proceedings of the Humanoids-2005 Workshop, 2005.

[11] Jane McGonigal, "*ENGAGEMENT ECONOMY: the future of massively scaled collaboration and participation*", September 2008 | SR-1183, INSTITUTE FOR THE FUTURE, 2008.

[12] Jane McGonigal, "*Reality Is Broken: Why Games Make Us Better and How They Can Change the World*", Penguin Press, 2011.

[13] Tatsuo Nakajima and Vili Lehdonvirta, "*Designing Motivation in Persuasive Ambient Mirrors*", Personal and Ubiquitous Computing, Vol. 17, No. 1, pp.107-126, 2013.

[14] Tatsuo Nakajima, Tetsuo Yamabe and Mizuki Sakamoto, "*Proactive Ambient Social Media for Supporting Human Decision Making*", In Proceedings of the 8th International Conference on Ubiquitous Intelligence and Computing, pp.25-39, 2011.

[15] Mizuki Sakamoto, Tatsuo Nakajima and Todorka Alexandrova, "*Digital-Physical Hybrid Design: Harmonizing the Real World and the Virtual World*", In Proceedings of the 7th International Conference on the Design & Semantics of Form & Movement, pp.211-222, 2012.

[16] Mizuki Sakamoto, Todorka Alexandrova and Tatsuo Nakajima "*Analyzing the Effects of Virtualizing and Augmenting Trading Card Game based on a Player's Personality*", In Proceedings of the 6th International Conference on Advances in Computer-Human Interaction, 2013.

[17] Mizuki Sakamoto and Tatsuo Nakajima, "*Value-based Design for Gamifying Daily Activities*", In Proceeding of the 11th International Conference on Entertainment Computing, pp.241-244, 2012.

[18] Dave Szulborski, "*This Is Not a Game: A Guide to Alternate Reality Gaming*", Lulu.Com 2005.

[19] Tetsuo Yamabe and Tatsuo Nakajima, "*Playful Training with Augmented Reality Games: Case Studies towards Reality-Oriented System Design*", Multimedia Tools and Applications, Vol. 62, No. 1, pp.259-286, 2013.

[20] Black★Rock Shooter, http://blackrockshooter.wikia.com/. (Octorber 12 2012)

[21] CyberOne TCG, http://www.c-1cyberone.com/official_top/official_top.html. (Octorber 12 2012)

[22] Duel Accelerator, http://www.yugioh-online.net/v3/newvisitors/. (Octorber 12 2012)

[23] The Eye of Judgment, http://en.wikipedia.org/wiki/The_Eye_of_Judgment. (Octorber 12 2012)

[24] Kyoto Prefecture and Ministry of Internal Affairs, "National Survey Promotion Video" , http://www.youtube.com/watch?v=IdAkKZKEfGU. (Octorber 12 2012)

[25] MikuMikuDance, http://www.geocities.jp/higuchuu4. (Octorber 12 2012)

[26] NFGD, "Guide Dogs Help Spread the Original Poster", http://www.nkoku.jp/pos. (Octorber 12 2012)

[27] The Legend of Zelda, http://zelda.com/. (Octorber 12 2012)

[28] Perplex City, http://perplexcitywiki.com/wiki/Main_Page

[29] Yu-Gi-Oh! Trading Card Game, http://www.yugioh-card.com/en/

[30] Institute of Government, *"MINDSPACE: Influencing Behavior through Public Policy"*, CabinetOffice, 2010.

An Interactive Agent Supporting First Meetings Based on Adaptive Entrainment Control

Tatsuya Hayamizu
Graduate School of Information Science and Technology
Osaka Institute of Technology
1-79-1 Kitayama, Hirakata City, Osaka, Japan
e-mail: hayamizu@is.oit.ac.jp

Mutsuo Sano
Facility of Information Science and Technology
Osaka Institute of Technology
1-79-1 Kitayama, Hirakata City, Osaka, Japan
e-mail: sano@is.oit.ac.jp

Kenzaburo Miyawaki
Facility of Information Science and Technology
Osaka Institute of Technology
1-79-1 Kitayama, Hirakata City, Osaka, Japan
e-mail: miyawaki@is.oit.ac.jp

Kentarou Mukai
Facility of Information Science and Technology
Osaka Institute of Technology
1-79-1 Kitayama, Hirakata City, Osaka, Japan
e-mail: m1m09a21@info.oit.ac.jp

***Abstract*—This paper describes an agent that can facilitate first meeting communications. In this situation, a communication mediator is important because people can feel stress and an inability to talk comfortably. Our agent reduces this stress using embodied entrainment and promoting communication. In previous research into embodied entrainment, appropriate back-channel feedback has been discussed but communication studies have been limited. We propose an embodied entrainment control system that recognizes a state of communication and is adaptive to each situation with effective nonverbal communication. In this way, our agent mediates a balanced, two-way conversation. Our experiments with the agent confirmed its effectiveness across various social skills levels. We demonstrate that the embodied entrainment of our agent in first meetings benefits people who have low social skills, thereby verifying the efficacy of our agent.**

Keywords-Embodied Entrainment; Nonverbal Communication; Introducer Agent; Group Communication Introduction; Social Skills

I. INTRODUCTION

Wider communication capabilities are required for robots to support humans' daily life. In today's world, it is not enough to have a simple relationship between a human and a robot; if robots are to support daily life, we need to develop true socialized robots.

We propose a robot design that can mediate a smooth group conversation and encourage human communication. Humans communicate with each other via many channels that can be categorized as verbal and nonverbal. Nonverbal communication channels such as intonation, accent, eye-contact, gestures such as nodding or other emotional expressions all encourage synchronization, embodied entrainment, and friendship [1]–[7]. Watanabe et al. [8] have investigated such a robotic agent, which includes embodied entrainment, and they developed a speech-driven interactive actor, the Inter Actor, which can predict appropriate timing of nods and gestures with voice information from a communication partner.

However, the Inter Actor focused on synchronization with a listener without regard to group communication that involves both the talker and listener in a balanced, two-way conversation. Kanda et al. [9], [10] adopted a joint gaze function for their robot and have investigated a more adaptive robot through a design-based approach. However, the research focused on rule-based natural behavior of robots rather than a robot design aiming to facilitate communication. In addition, the investigation did not mention group communication.

This paper focuses on group communication and presents a control method of embodied entrainment that is adaptive to the situation.

We begin with an introducer robot for communication between persons meeting for the first time (Fig. 1). Next, we explain our method for adaptive control of embodied entrainment and show the experimental results that verify the basic effectiveness of our agent. Finally, we analyze our agent's effectiveness from the viewpoint of social skills.

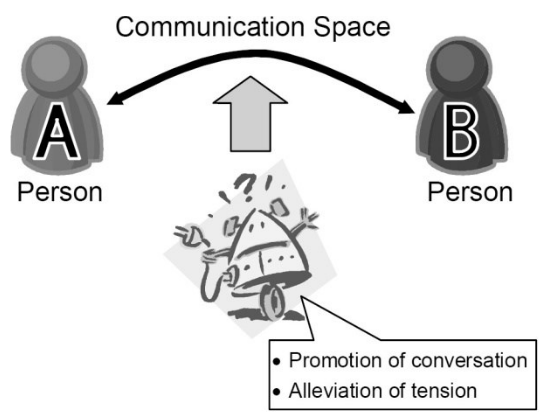


Figure 1. Introducer agent of a first meeting pair.

II. BASIC STRATEGIES TO PROMOTE CONVERSATIONS

This section describes the five mediating behaviors of our introducer agent to promote conversation.

 ISBN: 978-1-61208-250-9

A. Utterances inducement

The agent asks questions related to the participants in order to establish a mutual relationship between itself and each participant.

B. Gaze Leading

For transmission of information from participant A to participant B, the agent moves its gaze to A and asks him/her about the information. Then, the agent moves its gaze from A to B in order to achieve an equal eye-contact.

C. Gaze Distribution and Synchronizing Nod

When the participants talk to each other, the agent makes equal eye-contact with the participants and adjusts its nod timing to those of the participants'.

D. Dynamically Synchronizing

The synchronizing method is dynamically changed according to the state of the two participants as follows.

1) Natural synchronizing to a listener.

When a listener synchronizes with a talker by nodding and other back-channel feedback, the agent echoes these according to the listener's timing to encourage further communication.

2) Synchronizing to a talker to invite the other participant to the conversation.

When a listener does not listen to the talker, the agent tunes its back-channel feedback rhythm to that of the talker and makes eye-contact with the listener. By this behavior, the agent influences the listener and encourages flow of the conversation.

Fig. 2 shows the strategy details. Note that the symbols in the figure represent communication members as follows.

H1 Participant 1.
H2 Participant 2.
R Introducer agent.

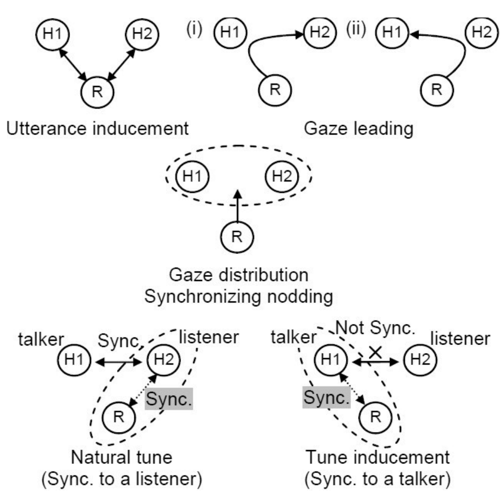


Figure 2. Introducer agent of a basic strategy.

III. ADAPTIVE CONTROL OF EMBODIED ENTRAINMENT

A. Macro Strategy to Promote Group Conversation

These behaviors must be executed in a conversational situation; therefore, we modeled the group communication on a first-time meeting and designed a macro strategy for the promotion of conversation. Fig. 3 describes the strategy and state transition model. The strategy consists of the grounding and enhancement processes on the top layer of the figure. The grounding process establishes a rapport and enhancement process promotes communications within the group. The bottom layer of Fig. 3 shows the state transition model. We segmented an introduction scene into five states on the basis of preliminary observations. The agent moves among these states according to the circumstances of the participants' communication as follows:

1) Greeting

The agent introduces itself and offers a brief explanation of the situation. Then, it introduces the participants' names.

2) Grounding

In this state, the agent tries to establish a relationship between itself and each participant. For this, the agent cites the participants' profiles and asks them simple questions. It aims to encourage them into the rhythm of the agent using utterance inducement. It is assumed that the agent has data such as name, hobbies, or other details of the participants who are being introduced.

3) Topic Offering

This state encourages conversation between the participants. The agent offers information and profiles of one participant to the other or asks a simple question in order to start a conversation between them. By such behavior, the agent manipulates the participants' gaze to make them communicate face to face.

4) Focusing on a Specified Topic

In this state, the agent tries to join in a conversation between the participants. The agent focuses on a topic that was offered at the previous state, Topic Offering.

5) Hearing Conversations

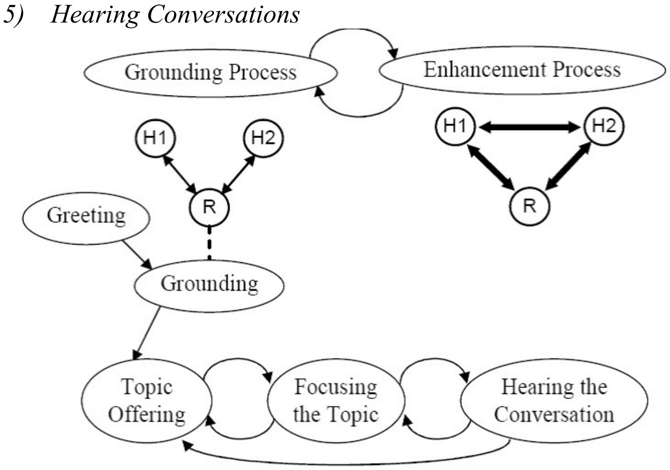


Figure 3. Macro strategy and state transition of an introduction scene.

After successfully making a close and friendly relationship between the participants, the agent keeps hearing their conversations quietly. In this state, the agent nods and looks at the talker using basic embodied entrainment strategies with correct timing.

B. *Communication Activity*

The agent monitors the participants' communication and estimates the current state as shown in Fig. 3. For the estimation, we define communication activity measurement. The agent calculates the activity in every time slice, defined by the average of speech level and mutual eye-contact of the participants. When this measurement satisfies the condition of a state transition, the agent moves to the next state. To explain this, we define the essential actions of communication by the following symbols that represent functions which return 1 when each action is detected by sensors. Otherwise, they return 0. Henceforth, the parameter t always denotes the time when the actions occur.

Nod (X, t)	X nods
Utterance (X, t)	X talks
Utterance (X rancewhen t	X talks to Y
Terminate Utterance (X, t)	X terminates his/her speech.
Gaze (X e/her sp	X directs his/her gaze to Y
Face (X irects h	X looks toward Y
Gaze (X ⇔ Y, t)	X's gaze accords to Y's
Face (X ⇔ Y, t)	X and Y look toward each other
Turn Gaze (R, X azeook	The agent guides X's gaze toward Y
Turn Utterance (R, X tter-ancee	R encourages X to talk to Y

Note that, in this paper:
Gaze (X → Y, t) = Face (X → Y, t) and
Gaze (X ⇔ Y, t) = Face (X ⇔ Y, t)

Our method calculates the communication activity using these symbols and is based on the following two equations:

1) *Ratio of gaze sharing*

$$R_{gazeShare}(x \leftrightarrow y, \Delta t) = \frac{\sum_{t=t0}^{t0+\Delta t} Gaze(x \leftrightarrow y, t)}{\Delta t}$$

2) *Average of utterance power*

$$Average_{Pu}(x, \Delta t) = \frac{\sum_{t=t0}^{t0+\Delta t} UtterancePower(x, t)}{\Delta t} + P_n$$

Note that, P_u (x, t) represents the power of x's utterance at the time t. P_n is a term of environmental noise. The following equation is an extended formula to calculate the average P_u of multiple participants.

$$Average_{Pu}(x, y, \Delta t) = \frac{\sum_{t=t0}^{t0+\Delta t} average\{P_u(x, t), P_u(y, t)\}}{\Delta t} + P_n$$

Note that Δt is a time span for calculation of the measurement.

Conditions of state transitions depend on the communication activity and are defined as follows.

1) *From Greeting, Grounding to Topic Offering*

In these states, the agent behaves according to the prescribed action scenario that is composed of general protocols of greeting and introduction.

2) *From Topic Offering to Focusing the Topic*

$$R_{gazeShare}(x \leftrightarrow y, \Delta t) \geq Threshold_{gaze}$$

3) *From Focusing the Topic to Hearing Conversation*

$$R_{gazeShare}(x \leftrightarrow y, \Delta t) \geq Threshold_{gaze}$$
and
$$Average_{Pu}(x, y, \Delta t) \geq Threshold_{utter}$$

In the equations, both the $threshold_{gaze}$ and $threshold_{utter}$ are acquired through learning.

C. *Learning for Adaptive Control of Embodied Entrainment*

The agent retrieves participants' information by sensors and determines its course of action to encourage entrainment. We built a decision tree (Fig. 4) of inductive learning for this interaction. The following three rules are extracted from the tree.

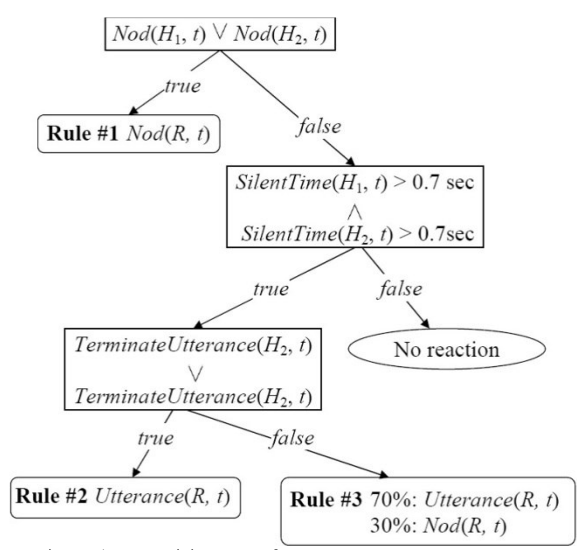


Figure 4. Decision tree of an agent.

1) Rule #1-Reactive nodding

This rule represents reactive nodding to a participant's nod.

2) Rule #2-An utterance

When one participant has finished speaking and the other participant stays quiet, it can be assumed that the conversation is over. The agent recognizes this situation and offers a topic to avoid any silence.

3) Rule #3-An utterance and nod to a participant's speech

This rule is very similar to Rule #2 but represents the situation where a participant stops speaking before finishing. In this situation, the agent is 70% likely to interject and 30% likely to nod. The criteria to determine whether a participant has finished speaking depends on the final part of the utterance. We extract the final part by Japanese morphological analysis.

IV. BASIC EFFECTIVENESS OF A FIRST MEETING INTRODUCER AGENT

Our aim is to promote exchange of information in introductory meetings and we conducted our experiment using an agent to mediate between strangers.

To validate the communication model in first-time meetings, one experiment explored communication between an agent and two strangers (condition with an agent), and the second experiment explored communication between two strangers (condition without an agent). To eliminate potential problems such as hearing difficulties, we excluded elderly participants from these experiments.

Fig. 5 shows our sensing environment allowing real-time gaze tracking of the strangers using multiple cameras (Figs. 5 (e)–(f)). A virtual agent is shown in the display used for this experiment ((Fig. 5 (g)). Fig. 6 shows our virtual agent. An acceleration sensor that was attached to participants detects gestures, whereas a microphone detects their speech (Fig. 7).

Experimental participants included 10 pairs of strangers, who were all university students.

From results of previous experiments, the time of this experiment was set to 6 min. In this experiment, topics offered by the agent included information about each stranger such as hobbies and special skills.

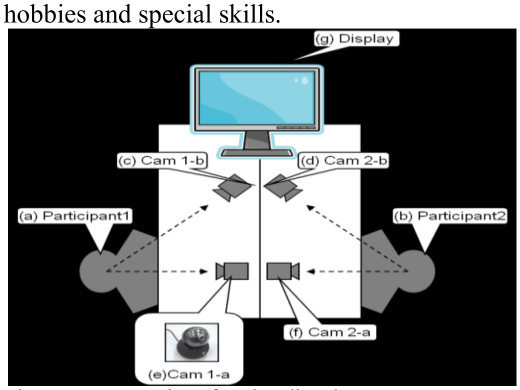


Figure 5. Detection of gazing direction.

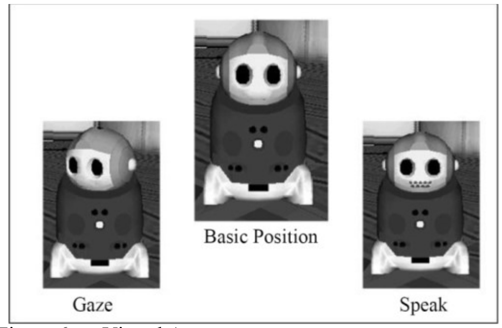


Figure 6. Virtual Agent.

Figure 7. Acceleration censor and microphone.

Fig. 8 shows an example of a communication field transition in this experiment.

Table 1 shows a statistical summary of the questionnaire that included questions such as "Was the conversation stimulating?", "Did you feel comfortable with your partner?", and "Did you come to know your partner well?". Answers were evaluated between 1 (the worst) and 5 (the best).

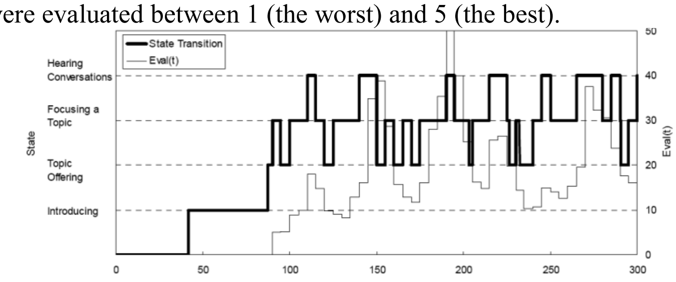


Figure 8. Communication field of state transition and activity.

TABLE 1. STATISTICAL SUMMARY OF QUESTIONNAIRE

Condition	Statistics	Is conversation excit	Did you feel your partner favor	Did you come to know your partner well ?
With Agent	Average	3.9	3.5	4.05
	Median	4	4	4.5
	Covariance	1.165287	1.192079	1.316894
Without Agent	Average	4.1	4.1	4.6
	Median	4	4	5
	Covariance	0.788069	0.788069	0.502625

We compared the questionnaire results of two experiment conditions, i.e., with and without an agent, using a t-test. The results of the t-test showed a higher score with an agent than without an agent particularly in regard to questions 2 and 3.

This validates the efficacy of our model for mediating first meetings.

V. Analysis of Agent from a Viewpoint of Social Skill

We examined the role of the agent's mediating behaviors with regard to participants' social skills, having already verified the basic effectiveness of the five mediating behaviors previously mentioned.

In this section, we hypothesize that embodied entrainment in mediating behaviors does not affect people with high social skills, but is effective for people with low social skills. Our experiments verify this.

People with low social skills often mistime speech, while people with high social skills risk monopolizing the conversation. We considered that our five mediating behaviors of utterance inducement, gaze leading, gaze distribution, synchronizing to a listener, and synchronizing to a talker are effective communication methods to defuse these situations for people with low social skills.

This experiment was conducted with pairs of strangers and an introducer agent and in conditions of with and without gaze sharing and nodding. In our previous experiment, the subjects were university students. In this test, we used elderly people in order to verify applicability of communication support in geriatric care as communication support for people with low social skills. The number of participants was 38 that included 18 elderly men and 20 elderly women.

The experimental environment was the same as the previous experiment, but this time, we used a real robotic agent for a more effective embodied entrainment than a virtual agent (Fig. 9).

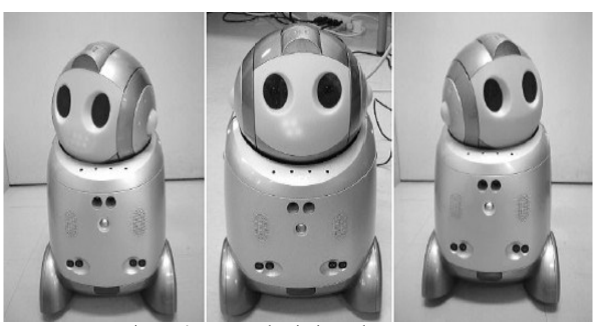


Figure 9. Robotic introducer agent.

We conducted a questionnaire before the experiment and inputted the information obtained to an introducer agent. We explain robot's mediating behaviors to two experimental participants and we shows questionnaire item in monitor to participants.

Each participant completed KiSS-18 for classification into groups of high and low social skills. Table 2 shows statistical summary of participants' social skills.

TABLE 2. Average of Participants' Social Skill

	Low	High
Sample	19	19
Average	57.2632	72.4737
Standard Deviatio	9.949	3.5492

Table 3 shows the national average of social skill [11].

TABLE 3. National Average of Social Skill

	National Average	
Man	61.82	(n=45,SD=9.41)
Woman	60.1	(n=121,SD=10.5)

After the experiment, the participants answered a questionnaire about the experiment with gestures and that without gestures. We conducted a t-test on the questionnaire with the results shown in Table 4.

TABLE 4. Results of the T-Test

Statistics	Low	High
t-Value	1.9509	1.2865
Degrees of Freedom	16	18
p-Value	0.0344(*)	0.1073

As shown by the t-test results, groups of high social skills cannot verify the efficacy of the agent either with or without gestures. However, groups with low social skills can verify the efficacy of an agent with and without gestures.

This result means that participants who have low social skills gain more from the five mediating behaviors.

VI. Conclusion

Our agent could change a state of communication such as Topic Offering and Focusing and Hearing on the basis of the communication activity measurement, which is calculated from mutual eye-contact and utterance power. On the basis of this state transition, we could stimulate communication through body entrainment behaviors such as nodding, eye-contact, and other back-channel feedback pertinent to each communication state. We conducted communication experiments between strangers and confirmed the effectiveness of adaptive state transitions and smooth interaction control. Our questionnaire confirmed that our agent could promote understanding between the communication pairs. From the evaluation, it is supposed that our system could deal with a stressful situation between strangers.

We aim to adapt our agent to promote information exchange between elderly strangers. For this, we conducted experiments in conditions of with and without gaze sharing and nodding.

From our results, we confirm the efficacy of our agent using gestures in groups with low social skills to encourage conversation between participants and demonstrate that the five mediating behaviors work effectively for participants with low social skills.

ACKNOWLEDGMENT

This research was supported by Mr. Ryouhei Sasama, Mr. Tomoharu Yamaguchi, and Mr. Keiji Yamada of NEC Corp. The authors would like to thank Enago (www.enago.jp) for the English language review.

REFERENCES

[1] J. D. Matarazzo, M. Weitman, G. Saslow, and A. N. Wiens, “Interviewer influence on durations of interviewee speech,” J. Verb. Learn. Verb. Beh., vol. 1, pp. 451–489, May 1963.

[2] J. T. Webb, Interview synchrony, “An investigation of two speech rate measures in an automated standardized interview,” in Studies in Dyadic Communication, B. Pope and A. W. Siegman (Eds.), New York: Pergamon, 1972, pp. 115–133.

[3] A. Kendon, “Movement coordination in social interaction: Some examples described,” Acta Psychologica 32:101–125, 1970.

[4] W. S. Condon and L. W. Sander, “Neonate movement is synchronized with adult speech: Interactional participation and language acquisition,” Science, vol. 183, pp. 99–101, January 1974.

[5] F. J. Bernieri, J. S. Reznick, and R. Rosenthal, “Synchrony, pseudosyn chrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions,” Journal of Personality and Social Psychology, vol. 54, pp. 243–253, 1988.

[6] Y. Takeuchiand and Y. Katagiri, “Emotional feedback by synthesized voice and facial expressions,” In Proccedings of the 3rd intenational conference of cognitive science, pp. 697–698, 2001.

[7] H. Kozima, C. Nakagawa, and H. Yano, “Can a robot empathize with people?,” International Journal of Artificial Life and Robotics, vol. 8, pp. 83–88, September 2004.

[8] T. Watanabe, “E-cosmic: embodied communication system for mind connection,” In Proc. of the 13th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN), pp. 1–6, 2004.

[9] T. Kanda, H. Ishiguro, T. Ono, M. Imai, T. Maeda, and R. Nakatsu, “Development of frobovief as platform of everyday-robot research,” The transactions of the Institute of Electronics, Information and Communication Engineers, J85-D-I(4):380–389, 2002.

[10] T. Kanda, H. Ishiguro, M. Imai, and T. Ono, “Development and evaluation of interactive humanoid robots,” Proceedings of the IEEE (Special issue on Human Interactive Robot for Psychological Enrichment), vol. 92, pp. 1839–1850, 2004.

[11] H. Hori and F. Yoshida, “Psychometric scale collection II,” Science, 2001.

The Virtual Counselor

Automated Character Animation for Ambient Assisted Living

Sascha Fagel, Andreas Hilbert
Research & Development
zoobe message entertainment GmbH
Berlin, Germany
{fagel|hilbert}@zoobe.com

Martin Morandell, Christopher Mayer
Health & Environment
AIT Austrian Institute of Technology GmbH
Vienna, Austria
{martin.morandell|christopher.mayer}@ait.ac.at

Abstract—We present a system for automated animation of text or voice messages suitable for Ambient Assisted Living user interfaces. Input to the system can be text, a pre-recorded speech file, or the speech signal captured directly from the microphone. Speech animation parameters are calculated by a co-articulation model either for the voice audio or – if available – from the phone chain extracted from the Text-To-Speech processing step in case of text input. An animation script that layers body movements and speech animation is generated. This script is rendered and converted into an h.264 video by a computer game engine. The system is developed to be used in care services for elderly users within a European research project.

Keywords: automatic character animation; embodied conversational agent; ambient assisted living; multimodal user interfaces; audiovisual speech synthesis

I. Introduction

Like in many other places in the world the average age of the inhabitants of the European Union is significantly increasing. This poses challenges to the society in various respects one of them being the need for more efficient healthcare for elderly people. Many current and future healthcare services will consist of a combination of human personnel and automated ict-systems, i.e., the service will at least partly be provided by a computer application.

Even though health- and homecare technology is often very complex in design, implementation and maintenance, the user interface towards the end user – in this case elderly people with all kinds of abilities, preferences and special needs – has to be kept very simple, easy to use and especially enjoyable and attractive. The user interface is the single component in such systems upon which everything else will be judged [1]. Therefore, in particular usability, accessibility, as well as the freedom of choice concerning the interaction with such systems are the crucial points for acceptability, applicability and subsequently the benefit of such systems – for the user him- or herself, for the society and also for the market-stakeholders.

The AALuis Project (Ambient Assisted Living user interfaces) [2,3] focusses on the aspect of freedom of choice for the preferred ways of user interaction. New approaches such as multi touch technologies and usage of avatars are developed and adopted to the very heterogeneous needs of primary end-users, elderly people who can derive a benefit from AAL Systems.

Embodied Conversational Agents (ECAs) that display appropriate non-verbal behavior were shown to enhance user satisfaction and engagement and improve the users' interaction with a computer system [4]. Therefore, the concept of an avatar that represents the service to the user as virtual personification was chosen as a core component of the AALuis-project user interface development. Furthermore, the addition of a visual display to verbal information – i.e., adding a lip-synched animated character to audio speech output – can increase the intelligibility and enhance the robustness of the information transmission [5] as known from natural speech [6].

This paper is structured as follows: The next chapter describes the animation system with its components animation generator, text-to-speech transformation, audio processing, and rendering. Chapter III describes how we embed the generator client and generated animations into user interfaces. We finish the paper with a discussion, and conclusions and future work.

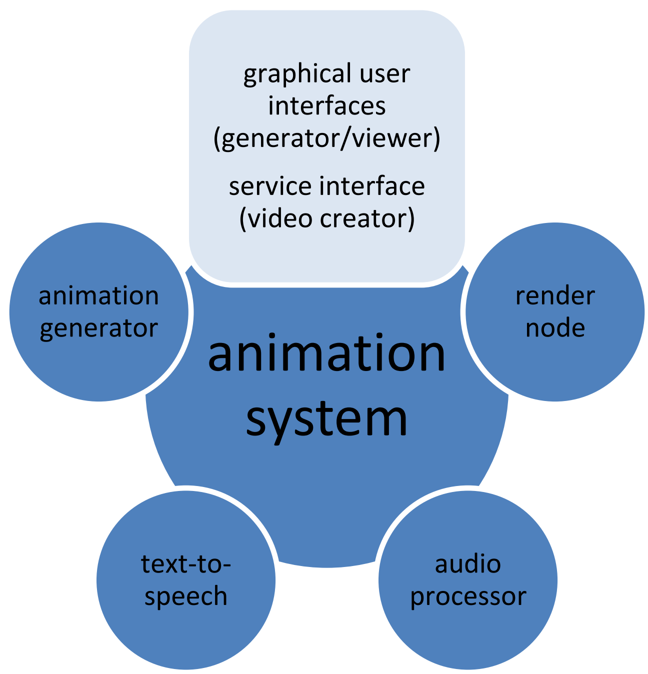


Figure 1. System overview.

II. Animation System

The core of the animation system is implemented in a JEE web application on a Red5 Media Server [7]. The web application provides a custom API that is accessible over the Real Time Messaging Protocol (RTMP, [8]). Besides the animation generation (see next section) the web application serves as a dispatcher of some sort – it manages the communication between the nodes that are necessary to fulfill the job. The data is exchanged between the components via TCP/IP. Figure 1 shows an overview of the system.

A Animation Generator

The animation generator collects data about the video that is to be created and generates an animation script accordingly. An animation script contains the following elements:

- the character model
- a sequence of animations for the character
- speech animation parameters
- the scenery
- the light set
- the camera settings
- the audio track
- technical data such as paths and encoder settings

References to all these elements (except speech animation parameters, the audio track and the technical data) and metadata such as the locations of the 3-D source files are stored in a database. The animation generator collects the necessary information and fills an xml data structure. The body animations are taken from a pool of animation cycles that are feasible to accompany speech and concatenated in a random sequence.

B Text-To-Speech

Data generated by different processing stages of the open-source-system MARY TTS [9] are used for Text-to-Speech-conversion. Alternatively, the high quality commercial CereProc cServer [10] is used. Animation parameters for lip-sync speech movements are derived from the duration generation stage which gives as output a chain of phonemes with their respective timings. Animation parameters are calculated for jaw opening, lip opening, lip spreading and tongue tip raising which are independently derived by an implementation of the dominance co-articulation principle [11]. Model parameters for ideal articulator positions and their dominances for a given phoneme are available from a study by Fagel and Clemens [12]. The presented method is modified in two details:

1. Instead of generating two target positions per phoneme – resulting in animation parameters not equally distributed over time – the animation parameters are determined based on equidistantly sampled phoneme values with a sample rate equal to the actual video frame rate.
2. The calculation is extended by a hypo/hyper-articulation parameter to generate slower movements with smaller magnitude or faster movements with greater magnitude, respectively. An activation value defined as “low”, “medium” or “high” that is assigned to the body animation entry in the database controls this articulation parameter.

C Audio Processing

Animation parameters for jaw opening, lip opening and lip spreading are calculated from a language independent, mostly rule-based audio analysis if speech recordings instead of TTS data are used. In a first step the audio signal is normalized, de-clicked, cropped and noise reduced.

Audio analysis for animation parameter extraction is carried out on the speech signal re-sampled to 22050Hz. Frequency-domain representations are gathered from that signal with an FFT and a step-size that equals or is greater than the video frame-rate with a minimum window overlap of 50%.

To determine the appropriate values for *lip spreading, jaw opening* and *lip opening* the following measurements are considered: broad-band energies narrow-band energies, tonality measure (1 - spectral flatness in the range 180Hz–1250Hz) and spectral difference (75Hz–3kHz) between two subsequent analysis frames (i.e., first order temporal deviation of the square-summed spectrum). All energy-based values are logarithmized. *Tongue tip raising* parameters are not calculated from the audio signal.

The narrow band energies cover the characteristic formant features of vowels which are then categorized as /a/-like, /e,i/-like and /o,u/-like sounds by different linear combinations of those measures and mainly help to determine the *lip spreading*. Those are later also weighted by the tonality measure which accounts for the amount of harmonic content in the signal and hence indicates vowels and voiced consonants.

Most influential for the *jaw opening* parameter are broad-band-energy measure, tonality measure and inversely the higher frequency measure accounting for sibilants.

The *lip opening* parameter is calculated solely based broad-band-energy measure and tonality measure.

Measurements are smoothed by a weighted moving average of 100ms.The weights of the measures are determined empirically and the parameters are normalized to range between 0 and 1. Final frame-wise animation parameters are extracted from the analysis parameter sequence by a band limited sinc interpolation.

D Rendering

The animation script is executed by a modified version of the open source game engine Nebula Device 3.0 [13]. We attached a server function to the engine in order to receive the render jobs and to deliver the results via TCP/IP. Furthermore, the frame sequence is captured from the video memory and passed to ffmpeg [14] for encoding. The final files are written to a network drive shared by the render server and the core animation system.

 ISBN: 978-1-61208-250-9

III. EMBEDDING THE ANIMATION

The animation system itself provides external interfaces for the generation and the delivery of animated speech messages. The user interface for the generation process either takes text as input, records a voice message from the microphone, or provides a selection of pre-recorded voice messages. This front-end is implemented in Flash ActionScript 3.0 [15] and uses the web browser's Flash plug-in as runtime environment. Figure 2 shows the text-to-avatar Flash interface.

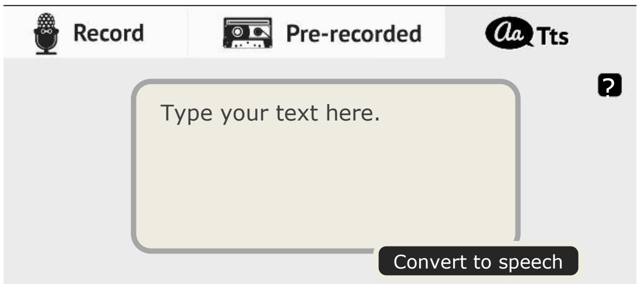


Figure 2. Text-to-avatar user interface.

Additionally to the (graphical) Flash video generation interface, a secure websocket based interface to create videos from text or speech audio is implemented in the OSGi framework [16] to create avatar videos from within any OSGi service. Character and scene settings are then defined by the service and given to the animation system by query parameters together with the text to be spoken or a reference to a (pre-) recorded speech audio.

When the video is rendered in the server back-end and encoded in h.264. This is delivered to the generating Flash application as preview, to a second Flash application (a simplified version of the one for generation), or can be embedded as a <video>-tag in an html5 website. The respective OSGi interface gets – as response to the call to create a video – a url for pseudo-streaming over http(s). Figure 3 shows a screenshot of the first prototype of the Virtual Counselor.

IV. DISCUSSION

There have been several approaches to use avatars in the field Ambient Assisted Living environments for elderly and personal home care assistants. Morandell et al [17] describe some advantages and a comparison of different types of avatars for this application field.

A very promising approach is the use of photo realistic avatars as personal assistants, in particular for the Human–Computer Interaction with elderly people with a diagnose of mild cognitive impairment. The presence of familiar faces can bring benefits such as higher acceptance, attention creation and creation of personal relation. A personalized non-photorealistic 3D avatar may have a comparable positive effect assuming it is clearly recognized as the intended person.

The study Avatars@Home [18] brought some insights concerning the use of Avatars compared to other output modalities. These findings are incorporated into the present approach that will combine a multimodal user interaction.

From the AALuis approach we expect the following advantages

- Joy of use: The high level of design and animation will lead to joyful use of the application. Combining entertainment, infotainment and edutainment.
- Broader applicability of the given avatars: The designed characters can be used for a broad field of information delivery and services.
- Fewer risks concerning personal relationships: Even though familiar faces could increase personal bindings this could also lead to interpersonal stress when an avatar represents a known person.

Interviews within the AALuis project on the possible usage of (personalized) avatars brought positive answers, in particular for applications such as tutoring. AALuis lab and field trials will bring deeper insights on acceptability, likeability and usability of avatars within Ambient Assisted Living environments.

V. CONCLUSIONS AND FUTURE WORK

We presented a system that generates videos of animated characters from speech or text. Although other applications are obvious, the animated character shown here is especially designed to serve as a virtual personification of an Ambient Assisted Living service in any user interface based on Adobe Flash or html5.

The communication between the front-end (user interface) and the back-end (animation system) via RTMP was recently extended by secure websocket connectivity to enable a wider variety of front-end solutions. iPhone and Android apps for mobile clients are currently being developed.

To ensure maximum platform independence of the user interface the animated voice message is delivered by displaying the accordingly generated video file. In the next version the avatar will be displayed on the screen with idle movements that fill the gaps between the informational outputs in order to represent the provided service persistently to the user. Methods for client-sided rendering of the avatar are currently under investigation.

Several male and female avatars will be modeled and animated in order to best represent the service and to best fit the user needs. The design of the avatars will be guided by a target group questionnaire.

Figure 3. The Virtual Counselor.

ACKNOWLEDGEMENTS

The presented work was supported by the European Commission under the Joint AAL Programme, the German Federal Ministry of Education and Research, funding reference No. 16SV5573K, the Austrian BMVIT and the programme "benefit".

REFERENCES

[1] van Berlo, A., "Design guidelines on smart homes", A COST 219bis Guidebook, 1999.

[2] AALuis project homepage: www.aaluis.eu [online resource, retrieved 18.10.2012]

[3] Mayer, C., Morandell, M., Hanke, S., Bobeth, J., Bosch, T., Fagel, S., Groot, M., Hackbarth, K., Marschitz, W., Schüler, C., Tuinenbreijer, K., "Ambient Assisted Living User Interfaces", Everyday Technology for Independence and Care - AAATE Assistive Technology Research Series 29, 456-463, 2011. DOI: 10.3233/978-1-60750-814-4-456

[4] Foster, M. E., "Enhancing Human-Computer Interaction with Embodied Conversational Agents", Proc. International Conference on Human-Computer Interaction, Beijing, 2007.

[5] Ouni, S., Cohen, M. M., Ishak, H. and Massaro, D. W., "Visual Contribution to Speech Perception: Measuring the Intelligibility of Animated Talking Heads", EURASIP Journal on Audio, Speech, and Music Processing, 2007.

[6] Sumby, W.H. and Pollack, I., "Visual Contribution to Speech Intelligibility in Noise", Journal of the Acoustical Society of America, 26, 212-215, 1954.

[7] Red5 Media Server 1.0, http://www.red5.org [online resource, retrieved 18.10.2012]

[8] RTMP Specification 1.0, Adobe Systems Incorporated, http://wwwimages.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/rtmp/pdf/rtmp_specification_1.0.pdf [online resource, retrieved 18.10.2012]

[9] Schröder, M. and Trouvain, J., "The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching", Intl. Journal of Speech Technology 6, 365-377, 2003.

[10] CereProc, cServer Text-to-Speech Server, http://www.cereproc.com/en/products/server [online resource, retrieved 18.10.2012]

[11] Löfqvist, A. "Speech as audible gestures", In W. J. Hardcastle and A.Marchal (Eds.): Speech Production and Speech Modeling, NATO ASI Series, 55, Kluwer, Dordrecht, 289–322, 1990.

[12] Fagel, S. and Clemens, C., "An Articulation Model for Audiovisual Speech Synthesis - Determination, Adjustment, Evaluation", Speech Communication 44, 141-154, 2004

[13] The Nebula Device. http://sourceforge.net/projects/nebuladevice

[14] FFmpeg. ffmpeg.org [online resource, retrieved 19.10.2012]

[15] ActionScript 3.0 Reference for the Adobe Flash Platform. http://help.adobe.com/en_US/FlashPlatform/reference/actionscript/3/ [online resource, retrieved 02.01.2013]

[16] The Open Services Gateway initiative framework. http://www.osgi.org [online resource, retrieved 18.10.2012]

[17] Morandell, M., Hochgatterer, A., Fagel, S. and Wassertheurer, S, "Avatars in Assistive Homes for the Elderly A User-Friendly Way of Interaction?", Lecture Notes in Computer Science 5298, 391-402, 2008. DOI: 10.1007/978-3-540-89350-9_27

[18] Morandell M, Hochgatterer A., Wöckl B., Dittenberger S. and Fagel S., "Avatars@Home InterFACEing the Smart Home for Elderly People", Lecture Notes in Computer Science 5889, 353-365, 2009. DOI: 10.1007/978-3-642-10308-7_25

AlgoPath's New Interface Helps You Find Your Way Through Common Algorithmic Mistakes

Estelle Perrin
CReSTIC
IFTS - University of Reims Champagne-Ardenne
Charleville-Mézières, France
estelle.perrin@univ-reims.fr

Sébastien Linck
IFTS - University of Reims Champagne-Ardenne
Charleville-Mézières, France
sebastien.linck@univ-reims.fr

***Abstract*—This paper presents the new interface of our serious game AlgoPath and its related interactions. AlgoPath helps students learn algorithmic. The virtual world represented in AlgoPath is all linked to the business of road construction and people running along these roads: objects students interact with are 3D figures, houses (huts and suburban houses), boxes, a crane, a concrete mixer and a bus station. This paper shows that AlgoPath helps students avoid common mistakes they can make while learning algorithmic. The entire interface is dedicated to help them conceptualize and understand the rules of algorithmic and programming. Whenever it is possible, AlgoPath reminds students of these rules and corrects the mistakes.**

***Keywords*-3D-based training; education; algorithmic; ludic teaching**

I. Introduction

When we first presented AlgoPath last year [1], we introduced the reasons why we had wanted it to be implemented: it was a necessity to have an entertainment computer program in which students could learn algorithmic but every single computer program we had looked into was simply an improved imitation of flowcharts. To renew interest, motivation, and enjoyment while learning, we had to achieve a virtual world in which students could create any algorithm they wanted. But to resemble video games, it had to be a world so we had spent time thinking of what a good concept of an algorithm could be. We had focused on how a variable should look like and had decided to turn it into a 3D white figure carrying a backpack that contains a value. In AlgoPath, a 3D white figure runs on a stone path. A path shape is related to the algorithmic statement it represents: linear when it is an assignment, forked when it is a conditional statement and circular when it is a loop.

During this year, we carried out a survey to show if students liked or disliked AlgoPath. Fifty students were quizzed. They came from different courses of studies - a half from under-graduate courses dedicated to websites; the other from scientific general under-graduate courses - but all had to attend a course dedicated to algorithmic. One hundred percent of students said they were thrilled to learn algorithmic with a game but sixty-five percent said the interface (see Figure 1) was neither ergonomic nor ludic.

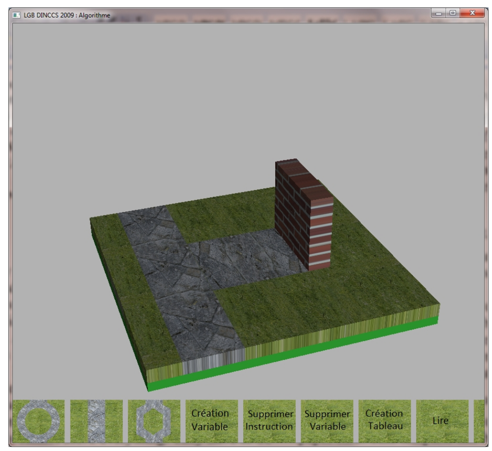


Figure 1. AlgoPath's old interface. Only the big buttons at the bottom can be selected.

They were right because AlgoPath was only a prototype implemented to validate the concepts, and we had not focused on the interface as we should have. So, the interface objects were succinct: ten big buttons displayed at the bottom of the window to create the basic statements of algorithms. Furthermore, to see the body statements of a loop or a conditional, students had to click on the corresponding bush along the stone path. Then the window completely changed and only showed statements of the loop or the conditional. Because the interface acted in that way, the students could no longer keep the whole concept in mind and were confused.

In this paper, we present the new interface of AlgoPath. This interface was designed (1) to avoid common mistakes students can do while learning how to create algorithms for the first time and (2) to navigate through a one and only world. The related works (see section II) show that a serious game interface must be close to a video game interface to be efficient. They also show that serious games can help to learn methods and rules proposed by a class. Section III is

 ISBN: 978-1-61208-250-9

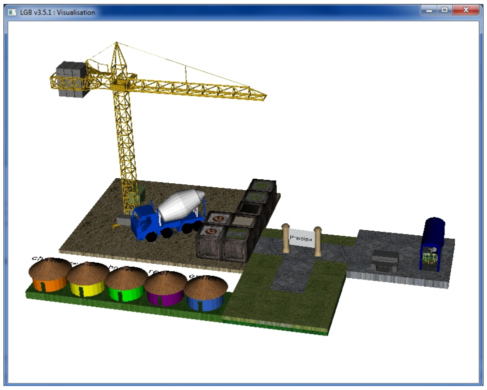


Figure 2. AlgoPath's new interface.

dedicated to the new interface of AlgoPath and the common mistakes it avoids. Section IV concludes the paper and shows how the evolution of AlgoPath might be.

II. Related Works

We can find many works on serious games. We will focus on three interesting characteristics of this kind of games: the creativity developed by the gamer, the graphical interface and environment of the game, and the usefulness of it.

A. Creativity

Serious games are made to learn something. Students can either acquire knowledge about courses studied at school or validate this knowledge with serious games.

Some are just made up of questions about the subject taught at school. A Quiz [2] or a maze with a range of issues related to a specific topic, [3] or [4], do not bring any part of creativity in the game. This first type of serious game has no creative part, it is more or less like school exams. Some of these games are labelled "serious games" only because of their simplified interface and their childish presentation, although it is no real support for learning whatsoever.

But on the other hand, some games develop a kind of interaction between the learner and the game, where the learner is involved in creating the game: for example series of actions by the player to create a solution to advance the storyline. The game does not only serve to check the acquisition of knowledge but actually allows the student to learn the concepts through play like: see [5] in physics, see [6] to learn problem-solving or see [7] in mechanical engineering. In [5] the authors show that the creativity increases when the player can share his creation with other players .

B. Interface

Serious games are first of all games, and so visual representation and interactions between the player and the interface of the game are very important. So, for people who play serious games, a game must be attractive in its form and in its way of learning. Today all the most played video games work in a 3D environment. If a serious game wants to be interesting for children or students, it has to be close to a classic game visual.

Questioning games are classically in only two dimensions. But some of them use 3D interface to interact with the player. These 3D-games are often made in 3D because of the subject of the game: geography [8] or architecture [9], but some are just made in 3D to have a better interface for collaboration between players [10]. Some serious games are just serious scenarii based on a commercially available video games [11], [12]. For example in [12], they use Tycoon City: New York® and SimCity Societies® games to learn daily economics and global issues.

But all these 3D interfaces and interactions do not frame the learning.

C. Usefulness

Many papers have been written about the utility of games in education. Whatever the age of the students, the learning with serious game is equivalent to that done in a master course[13]. For [14], serious games are no obstacle to students' success. From primary school to high school, the enjoyment to play [15] and, so to learn, is a good indicator of the usefulness of serious games.

To be effective, a serious game must be a game as well as a teaching aid. In that respect, teachers must be given a special training to be able to use serious games properly[16]. Then, if both teachers and learners use the game in the right way, then the game itself will gain in efficiency as well as in usefulness.

We have presented many serious games, some of which allow the students to get interested in discovering ans learning new subjects. We will next show how the new interface of AlgoPath promotes the learning of algorithmic.

III. AlgoPath's new interface

AlgoPath's interface was totally rethought in order to simplify the navigation and to help students avoid most of the common mistakes they can make while they learn algorithmic. The world within witch students can play is divided into five zones. The first one is the stone path. This zone represents the algorithm being built and it evolves gradually when students interact with the environment. The second one is a group of five houses. While interacting in this zone, students can create variables. AlgoPath provides five basic data types: integers, floats, Booleans, strings and characters. The third zone is the construction zone. Students can create aggregate or composite data type, used to represent entities that are described by multiple attributes of potentially different types. Students can create statements and they can build the prototype of a function or a procedure. The fourth zone is the bus station. This zone is dedicated to the simulation of the execution of the algorithm. And the fifth zone is the variables area. The latter is missing in Figure 2. because it is related to the memory usage and at the beginning there is no variable declared.

In the next sections, we describe the last four areas in details. The first one (the algorithm zone) was fully described in our previous paper [1] but, as the bush statements world totally replaced the main statements world when the user wanted to focus on the bush statements, we include a section in this paper in which we explain how we changed that.

A. Algorithm Zone

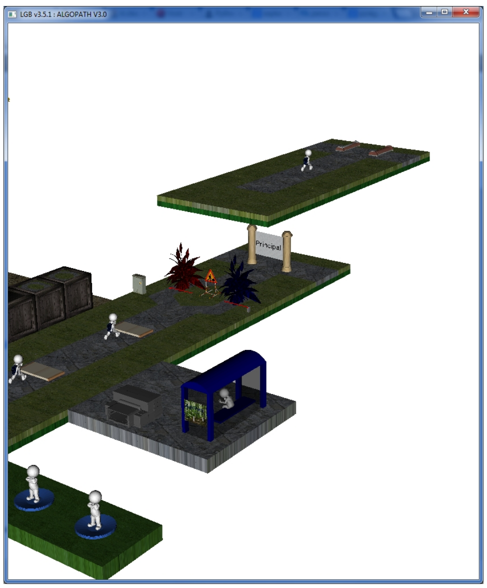


Figure 3. Level 0 and level 1 open.

In AlgoPath, a bush is a sequence of statements that describes actions to be performed. Replacing a world with another when students clicked on a bush to see its relative statements was not satisfying. Students could not locate themselves in the levels of decomposition of the algorithm. In this new version of AlgoPath, whenever students want to focus on statements of an else-part, a then-part, a loop, a function or a procedure, AlgoPath adds a new floor above the mother statement (see Figure 3). In that way, students do keep in mind the all concept. But, since a conditional statement can lead to two sequences of statements of the same level, AlgoPath does not allow to open more than one floor at a given level. Therefore, each opened floor of a superior level has to be closed before another floor can be opened.

B. Urban Zone

In the urban zone (see Figure 4) AlgoPath provides five basic data types: integers, floats, Booleans, strings and characters. Each data type has its own hut. The population

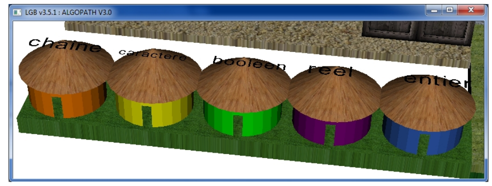


Figure 4. Urban zone.

of integer variables lives in the integer hut. The name of the data type is on the top of the hut and can be easily seen by students. Students can click on a hut as they would knock at a door and then trigger a variable. Clicking on a hut is one of the only few things students can do at the beginning of a session with AlgoPath (apart from creating a new composite data type; creating the prototype of a function or a procedure; and adding an output statement).

AlgoPath requires a name to declare a variable. If the name is the same as the one of an existing variable, AlgoPath declines the declaration and alerts students a variable has already been declared with this name. Just like in programming, words separated by a space are not accepted.

C. Memory Zone

The memory zone shows each declared variable. As seen in [1], a standing 3D figure personifies a variable. In this new version it stands on a pedestal whose colour is the same as the roof of the hut the 3D figure belongs to. A 3D red figure means the variable has not been assigned a value yet.

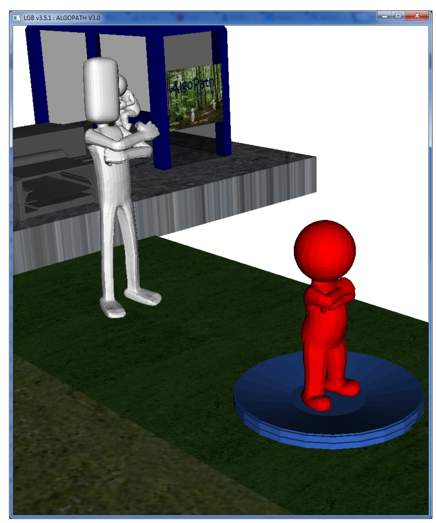


Figure 5. Two kinds of 3D figures.

Two kinds of 3D figures can stand on the memory zone (see Figure 5). A small and a little overweight one represents a basic data type variable; on the contrary, a tall and thin one represents an aggregate or composite data type variable. If students click on a tall and thin 3D figure, it opens a floor above it, with the 3D figures corresponding to the members

 ISBN: 978-1-61208-250-9

of the composite or aggregate type, just as a new floor opens when students click on a bush.

D. Construction Zone

In the construction zone, students can find boxes (to create statements), a concrete mixer (to create new prototypes of functions or procedures), a crane (to create an aggregate data type), and a dustbin (to delete things). Each object is dedicated to an interaction described in the following sections.

1) Concrete mixer: If students click on the concrete mixer, Algopath helps them go through the creation of the prototype of a function or a procedure. When writing algorithm on a sheet of paper, students often forget syntax or are embarrassed deciding if they have to write a procedure or a function. According to our rules, if a module has only arguments by value and one result then it is a function. In any other case, we ask students to write a procedure (meaning a mix of arguments by value and by reference with zero or at least two results). First AlgoPath asks them if it is a function or a procedure that will be created. The name of the module is then required and AlgoPath launches the process of creating arguments. An argument is totally defined by its name and its type. AlgoPath lets students choosing the name but it automatically suggests the set of the available data types. So, students cannot specify a data type they have not already defined. But when you define a procedure in a programming language, you have two choices regarding how arguments are passed to it: by reference or by value. In AlgoPath, you have three choices: by input, by output, and by input-output. Passing by input refers to a way of passing arguments where the value of an argument in the calling function cannot be modified in the called function. Passing by output refers to a way where an argument has no value in the calling function prior to the called function, but the called function has to assign it a value. Passing by input-output refers to a way of passing arguments where the value of an argument in the calling function can be modified in the called function. There is a distinction between argument by input-output and argument by output because we want students to be aware that an argument by reference may not be assigned a value when it reaches the called procedure. If students want to create a procedure with zero or several arguments by input and only one argument by output then AlgoPath warns them that it should be a function instead of a procedure and adds it as a function.

Once the prototype of the module is defined, a new box appears in the construction zone. This new box is selectable. If students select it then a new floor opens in AlgoPath. It shows the - empty - body of the module. Next to the stone path where statements will be added, there is a zone that looks like the memory zone of the main algorithm. Instead of showing the main variables, it shows the arguments of the module. Visible features (see Figure 6) help the students

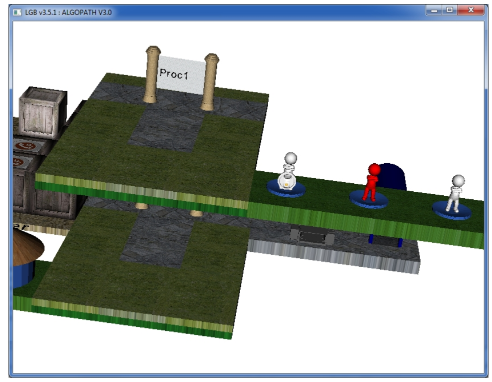


Figure 6. Three kinds of arguments (from left to right: by input, by output and by input-output).

recognize if it is an argument by input, by output or by input-output. An argument by input is a 3D white figure in front of which is a padlock. It means this variable was assigned a value and this value cannot change. An argument by output is a 3D red figure. It means it has not been assigned a value yet. An argument by input-output is a 3D white figure without a padlock. It means it has already been assigned a value and this value can change.

When students add an assignment statement in the body of a procedure, AlgoPath suggests them the set of variables available. Arguments by input are not included in this set. It is a little bit restrictive regarding programming, but it helps them understand that an argument by input cannot change its value.

2) Crane: Clicking on the crane launches the creation of composite data types. Students can create arrays or structures. An array is a set of consecutive variables of a same type. A structure is a collection of one or more variables, possibly of different types, grouped together under a single name for convenient handling. The variables named in a structure are usually called members. Students have to specify the types of cells of an array and the members of a structure. AlgoPath helps them with this process by suggesting the data types available. That is how AlgoPath avoids common typing mistakes made by students - writing a data type that does not exist or wrongfully writing a data type that actually exists. But it also helps the students understand that the most inner data types must be created first. If a person structure has to be created - which we will assume has a name and a date of birth - the date structure has to be created first. Naming an array, a structure and its members is another process students have to go through. AlgoPath verifies the integrity of the names. Naturally, a structure member and an ordinary (i.e., non-member) variable can have the same name without conflict, since they can always be distinguished by context.

3) Boxes: The construction zone contains seven boxes. Each box creates a new statement in the path of AlgoPath.

 ISBN: 978-1-61208-250-9

Figure 7. The seven boxes to create statements.

Figure 7 shows the seven boxes. The lower row contains - from left to right - a box to create an assignment, a box to create an input statement, a box to create an output statement, a box to create a conditional statement and a box to create a loop. Since these five statements were fully described in [1], this section will focus on the last two boxes: the one with a receiver and a ”F” and the one with a receiver and a ”P”. But let’s just add that anytime the name of a variable is required, AlgoPath suggests a set of proper names available. To fully explain AlgoPath does not only check the names of the variables declared, it also computes the names of the variables of basic data types because one of our rules is that a variable of aggregate data type cannot be assigned a value. For example, if the variable P is declared as a person, that is a structure with two members - a name and a date of birth - whose names are ”name” and ”date” and, if a date is a structure with three members named month, day and year then AlgoPath suggests the following set of names: P.name, P.date.month, P.date.day and P.date.year. P and P.date do not belong to this set because they are composite data type variables.

The upper row left box creates a function calling statement, while the right box creates a procedure calling statement. These creations can occur within two contexts. The prototype of the function or the procedure may be already defined or not. If it is, AlgoPath helps students by notifying the arguments of the module and their status: by input, by output and by input-output. For each, it reminds students if they can associate a value or a variable of the calling module. Naturally, AlgoPath acts differently whether an argument of the module is exclusively by input or not. If it is, AlgoPath lets students choose if they will associate a value - an expression - or a variable. If it is not, AlgoPath only shows variables of the same data type of the calling module. If there is none, AlgoPath cancels the process of creating a new statement and explains why. Then students learn they first have to create variables in the calling module if they want to put results in them. If the prototype is not defined, AlgoPath asks questions so the students are able to define the calling arguments of the module. First, it asks what the name of the module is. Then it asks if students want to add an argument. If they do, it wants to know if the argument is a value or a variable. In case it is a value, the argument is automatically defined by input. In case it is a variable, students can choose if it is by input, by output or by input-output. At the end of the process, the prototype of the module is automatically created and a new box is added in the construction zone.

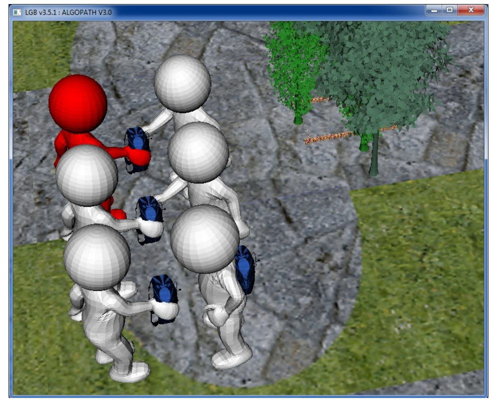


Figure 8. Visible features of calling arguments (colors and backpacks).

As mentioned in [1], there are visible features to tell the arguments apart. But in this version of AlgoPath, a calling argument by output is red while a calling argument by input-output is white (see Figure 8).

4) Dustbin: In the construction zone, there is a large commercial refuse bin. It opens when students click on it if a statement has already been added. Otherwise it does nothing. It closes when students choose the statement they want to delete.

E. Bus Station Zone

The bus station zone is under construction so we won’t talk much about it. This zone is dedicated to the execution of an algorithm. We expect a lot from this zone because normally students have to wait for the implementation of algorithms in a chosen programming language to discover the execution. Let’s just say a bus will drive along the path of AlgoPath to pick up the 3D figures (see Figure 9).

Figure 9. Execution.

 ISBN: 978-1-61208-250-9

The interface will act accordingly: for example, gates will open or close, 3D figures will tell the content of their backpacks, etc. Moreover, when students click on the printer, AlgoPath shows them the algorithm writing with conventional declarations and statements.

IV. Conclusion and Future Work

[1] presented the different 3D objects to represent the concepts of algorithms. A survey showed that students were not satisfied with the interface. The latter was too dour and even though students were able to avoid some grammatical mistakes, the computer program did not prevent them all and did not teach students why they were about to make mistakes. The new version of AlgoPath now implements those features. With AlgoPath, students can no longer:

- Add a statement if it is not an output statement before a variable is declared;
- Add a variable in a R-value of an assignment if it was neither declared nor assigned a value;
- Assign a value to an aggregate or composite data type variable;
- Try to assign a variable if it was not declared;
- Forget or add the calling parameters of a module: the number of calling parameters are always equal to the parameters of the module;
- Associate a parameter of a prototype with a calling parameter of a different type;
- Use a composite data type if it is not declared;
- Forget what the basic data types are;
- Declare a variable if AlgoPath does not know the type;
- Forget what statements they can use;
- Assign a value that does not match the type of the variable.

Moreover, when students are about to make a mistake and AlgoPath is able to notice it, AlgoPath explains why it is a mistake and corrects it. In the future, we will focus on adding Object-oriented programming concepts such as objects, classes, data abstraction, encapsulation, polymorphism, and inheritance. We will also study the possibilities to add distributed algorithms and programming concepts.

Acknowledgment

We thank Dimitry Zekrouf, Nicolas Fleurentin, and Sonny Jestin for their valuable help during the study and the implementation of this new version of AlgoPath.

References

[1] E. Perrin, S. Linck, and F. Danesi, “Algopath: A new way of learning algorithmic,” in *The Fifth International Conference on Advances in Computer-Human Interactions*, Valencia, Spain, 2012.

[2] W. Barendregt and T. M. Bekker, “The influence of the level of free-choice learning activities on the use of an educational computer game,” *Computers and Education*, vol. 56, no. 1, pp. 80–90, 2011.

[3] M. Virvou and G. Katsionis, “On the usability and likeability of virtual reality games for education: The case of vr-engage,” *Computers and Education*, vol. 50, no. 1, pp. 154–178, 2008.

[4] M. Papastergiou, “Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation,” *Computers and Education*, vol. 52, no. 1, pp. 1–12, 2009.

[5] D. B. Clark, B. C. Nelson, H.-Y. Chang, M. Martinez-Garza, K. Slack, and C. M. D’Angelo, “Exploring newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in taiwan and the united states,” *Computers and Education*, vol. 57, no. 3, pp. 2178–2195, 2011.

[6] C.-C. Liu, Y.-B. Cheng, and C.-W. Huang, “The effect of simulation games on the learning of computational problem solving,” *Computers and Education*, vol. 57, no. 3, pp. 1907–1918, 2011.

[7] B. Coller and M. Scott, “Effectiveness of using a video game to teach a course in mechanical engineering,” *Computers and Education*, vol. 53, no. 3, pp. 900–912, 2009.

[8] H. Tuzun, M. Yilmaz-Soylu, T. Karakus, Y. Inal, and G. Kizilkaya, “The effects of computer games on primary school students’ achievement and motivation in geography learning,” *Computers and Education*, vol. 52, no. 1, pp. 68–77, 2009.

[9] W. Yan, C. Culp, and R. Graf, “Integrating bim and gaming for real-time interactive architectural visualization,” *Automation in Construction*, vol. 20, no. 4, pp. 446–458, 2011.

[10] R. Hamalainen, “Designing and evaluating collaboration in a virtual game environment for vocational learning,” *Computers and Education*, vol. 50, no. 1, pp. 98–109, 2008.

[11] D. Charsky and W. Ressler, “”games are made for fun”: Lessons on the effects of concept maps in the classroom use of computer games,” *Computers and Education*, vol. 56, no. 3, pp. 604–615, 2011.

[12] Y.-T. C. Yang, “Building virtual cities, inspiring intelligent citizens: Digital games for developing students’ problem solving and learning motivation,” *Computers and Education*, vol. 59, no. 2, pp. 365–377, 2012.

[13] M. Ebner and A. Holzinger, “Successful implementation of user-centered game based learning in higher education: An example from civil engineering,” *Computers and Education*, vol. 49, no. 3, pp. 873–890, 2007.

[14] L. A. Annetta, J. Minogue, S. Y. Holmes, and M.-T. Cheng, “Investigating the impact of video games on high school students’ engagement and learning about genetics,” *Computers and Education*, vol. 53, no. 1, pp. 74–85, 2009.

[15] F.-L. Fu, R.-C. Su, and S.-C. Yu, “Egameflow: A scale to measure learners’ enjoyment of e-learning games,” *Computers and Education*, vol. 52, no. 1, pp. 101–112, 2009.

[16] D. J. Ketelhut and C. C. Schifter, “Teachers and game-based learning: Improving understanding of how to increase efficacy of adoption,” *Computers and Education*, vol. 56, no. 2, pp. 539–546, 2011.

A Three-Dimensional Interactive Simulated-Globe System Application in Education

Wei-Kai Liou
Science Education Center and Graduate Institute of Science Education, National Taiwan Normal University, Taipei, Taiwan(R.O.C)
Email: liouweik@ntnu.edu.tw

Chun-Yen Chang
Science Education Center and Graduate Institute of Science Education and Department of Earth Sciences, National Taiwan Normal University, Taipei, Taiwan(R.O.C)
Email: changcy@ntnu.edu.tw

Abstract

This study proposes an innovative three-dimensional (3D) Interactive Simulated-Globe System application. The instrument includes: a data processing unit, a wireless control unit, an image capturing unit, a laser emission unit as well as 3D hemispheric body imaging. The 3D hemispheric body imaging is designed to display the output image from a data processing unit. The Laser emission unit is for emitting a laser spot on the output image. Based on the spherical coordinates of the laser spot, detected by a data processing unit through an image capture unit, the spherical coordinates, with a plane coordinates converter operation, can provide spherical coordinates to convert into plane coordinates for the data processing unit. Utilizing the coordinates location, of the laser spot through image acquisition, calibration and internal coordinates with external coordinate converter operation can be guided the cursor, output by the data processing unit on 3D hemispheric body to launch synchronous movement with laser spot. To combine wireless control technology can perform synchronous driven, control and interactive with the software image on 3D hemispheric body. This work allows general planet software such as Google Earth (Mar, Moon), used in a variety of panel displays or projection turn into the 3D hemispheric body, and then through the laser spot emitted by the laser emission unit and wireless control unit, allows the laser spot in synchronization to control the cursor and software on the 3D hemispheric body imaging for a variety of interactive, and then perform the effect of 3D interactive globe system in any classroom or astronomical Museum for formal and Social education.

Keywords-three-dimensional; interactive; spherical coordinates; internal coordinates; external coordinate

I. INTRODUCTION

The various sizes of earth or planet instrument models have been widely used in traditional classroom and astronomical museums; they have become an integral part of teaching aids for formal or informal education teaching field. [1-3] In recent years, related inventions have attempted to use projection to project earth or other planets software images as digital teaching auxiliary aids. [4-6] However, the way to project earth or other planets software image on the screen is not much difference to use traditional earth or planet instrument for the teacher in the teaching and can not to fully demonstrate the advantages of digitization teaching aids in this way. Therefore, we propose a kind of earth or planet instrument, which can take advantage of simple technology and classroom existing equipment. This instrument is employed to project the earth or other planets images as three-dimensional sphere surfaces. By forming 3D earth or plants images, and utilizing the interactive technology of this work, the operator can use a simply handheld wireless control device to interact and synchronize control the cursor and function then perform the effect of an 3D interactive globe system.

II. RELATED WORK

Existing applications of laser guidance technology, U.S. Patent No. 6,275,214 [7], use a variety of different light sources for projection to projection curtain via the camera receives images. Through image processing and computer software program calculated the light projected onto the screen spot position of the projection screen to convert the signal to control a computer mouse cursor moves. The existing the whiteboard also uses the image capture element to achieve method (internal coordinates based on the coordinates of the image input is converted to the coordinates of the external programs and methods), in U.S. Patent No. 4,507,557 [8]. At present, the 3D image projection method synchronizes the projector in a different direction and position on the respective image in a 3D sphere. The corresponding image project on a spherical surface according to each specific projector to multiple projectors, each represents a specific image region of the sphere to generate a complete surface projection mapping imaging such as U.S. Patent No.0,256,302 [9]. Other method as 3D internal back projection methods, this method is to use the projector image, projected to an inner surface of the 3D projection surface. The formation of the 3D projection surface outside observer can see the imaging. Application examples are celestial bodies such as the earth, the planets and the anatomical organs in detail, such as the U.S. Patent No. 8,066,378 [10]. A 3D convex surface of the

 ISBN: 978-1-61208-250-9

display system is the system's display surface having a 3D convex shape. Projector system sucked image output to the internal display surface of the display surface, the image due to the convex shape of the display so that the viewer can see the 3D stereo images. For details is on U.S. Patent No. 7,352,340 [11]. The 3D internal projection system is to use optical scanning or projection, projected on a large 3D objects on the inner surface, so that when viewed by an external stereoscopic imaging effect, such as the U.S. Patent No. 0,027,622 [12].

III. PROPOSED THEOREM

This research main technology is divided into two parts. The first part is laser guide technology to control cursor. For controlling the cursor directional control function, it needs to install a laser emitter source and utilizes an optical camera with special light filter lens receiver to detect the laser source on the screen. After some image process and the spherical coordinates with plane coordinates converter operation, the movement of image can be transformed by a cursor command to guide the cursor directional control function. The second part is to install a wireless control electron circuit in our system. By linking the wireless control technology, it can direct execute each kind of interactive control function completely.

A. *Image acquisition*

The 3D hemispheric body imaging is to display the output image from data processing unit as shown in Figure 1. Laser emission unit is for emitting laser spot imaging on hemispheric body and the spherical coordinates of laser spot was detected by a data processing unit through the image capture unit. The key point of image acquisition technology in this study is the red visible light filter when an optics camera is used. This kind of special light filter lens may filter the Blue and Green visible light on the 3D hemispheric body imaging effectively, and causes optics photography function effective to trace the laser emitter spot on the three-dimensional imaging hemispheric body. The wherefores is red visible light filter lens make whole image and background turn into full red view. And it is because the wavelength of red laser spot is higher than any visible red color. When using this spot tracking can provide spherical coordinates of laser spot precisely.

B. *The spherical coordinates with plane coordinates converter operation*

Laser emission unit is for emitting laser spot imaging on hemispheric body. Based on the spherical coordinates of laser spot, detected by a data processing unit through the image capture unit and the spherical coordinates with plane coordinates converter operation can make spherical coordinates convert into plane coordinates for the data processing unit. By the coordinates location of the laser spot through image acquisition, calibration and internal coordinates with external coordinate converter operation can be guided the cursor, output by the data processing unit on 3D hemispheric body imaging to launch synchronous movement with laser spot as shown in Figure 2. The formula of spherical coordinates convert into plane coordinates as follow:

$$r = \sqrt{x^2 + y^2 + z^2}$$

$$\theta = \arctan\left(\frac{\sqrt{x^2+y^2}}{z}\right) = \arccos\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right)$$

$$\phi = \arctan\left(\frac{y}{x}\right)$$

On the contrary, formula of plane coordinates convert into spherical coordinates as follow:

$$x = r\sin\theta\cos\phi$$
$$y = r\sin\theta\sin\phi$$
$$z = r\cos\theta$$

C. *Wireless control technology*

By wireless control technology, this work can perform control and interactive with the software image on 3D hemispheric body imaging. We have succeeded to develop a useful technique for the classroom and museum, and have obtained the solution of the man-machine interface for its interactive operations. A three-dimensional interactive globe system was proposed in this study. This invention makes general planet software, used in a variety of panel displays or projection turn into the 3D hemispheric body imaging. Through the laser spot emitted by the laser emission unit and wireless control unit, allows the laser spot to synchronize control the cursor and software on the 3D hemispheric body imaging for a variety of interactive. Then perform the effect of 3D interactive globe system in education.

IV. IMPLEMENTATION

A. *System installation*

Figure 3 shows the schematic diagram of 3D interactive simulated-globe system. In this figure, the system contains information processing unit, image capture Unit, a three-dimensional imaging hemispheric body, laser emitting unit and wireless control unit. The 3D hemispheric body imaging is to display the output image from data processing unit. The operator held a laser emission unit for emitting

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-250-9

laser spot on imaging hemispheric body in front of hemispheric. The laser spot position was detected by a data processing unit through the image capture unit. By program operation can drive the control cursor on 3D hemispheric body imaging to execute simultaneous movement with laser spot. Wireless control unit is a device, installed a wireless control electron circuit inside. Each button has unique command function for remote control information processing unit then can perform control command function. By simple technology and classroom existing equipment, this system shows an effect solution for interactive of three-dimensional interactive globe system.

B. Calibration

The data processing unit by the image capture unit to detect five laser spots on the 3D hemispheric body imaging in sequence at initial. Based on the five laser spot localization method by internal plant coordinates with external spherical coordinates conversion program operator can calibrate the coordinate of control cursor on the imaging hemispheric body at the outset. Every point for calibration takes approximately a few seconds for each. That is followed by a defined sequence for completing five points calibration. When each point is captured by the camera the tracking process can be started, the cursor tracking is set up completely. By this five laser spot localization method, the control cursor on the imaging sphere can launch synchronous movement with laser spot precisely.

C. Interactive control

To combine laser guide technology and wireless control technology, this work can fully demonstrate the advantages of digitization 3D interactive simulated-globe teaching aids in this study. For example: the operator can rotate the earth or the planet, drag, draw, click, move, zoom in and zoom out any icon or function to directly interactive with 3D simulated-globe as show in Figure 4. Whenever to control the laser spot, stopped at the desired tap objects imaging on 3D simulated-globe then continuously pressing left mouse button function key on handheld device and we can perform a drag function in this system as we want. If we want to rotate this 3D simulated-globe, we only have to control the laser spot in any position on the globe and to press left mouse button function key on handheld device and quick move to left or right then release the button at the same time. The 3D simulated-globe will continuous rotate following by our movement speed and direction until the laser spot on globe and to press left mouse button function key again to stop it.

As we want to select any objects or icons on the 3D simulated-globe, the laser spot pointed objects or icons and click left mouse button function key on handheld device and the objects or icons will be selected. The zoom function in our system is to point the laser spot on the desired imaging objects and click left mouse button function key on handheld device to select an objects. Then subsequently press the zoom in or zoom out function button on handheld device issued the command to control zoom the desired imaging objects on 3D simulated-globe. Based on zoom function in this study, the pointed object by the laser spot on globe will be enlarged or reduced as we want. Such as when forward mouse wheel or key in PgUp is making a selected object to enlarge. On the contrary, backward mouse wheel or key in PgDn is making the selected objects to shrink. With drag and click function as mention above, we also can use drawing software to draw any sign, mark, emblem, label and symbol on the 3D simulated-globe.

The operation of the present study described as follow: First step, the data processing unit output the image on hemispheric body by a projector as shown in Figure 4. Subsequently, an operator held a device, integrated by the laser emitting unit and the wireless control unit. After, the operator points to the laser spot on an imaging hemispheric body. Then, the coordinate position of laser spot on the imaging hemispheric body was detected through the image capture unit. At the same time, data processing unit executed the internal coordinates with external coordinate converter operation and calibration. By laser-guided control program and the five points calibration program, enables the control cursor on the imaging hemispheric body can be synchronized movement with laser spot. With conjunction laser emission and wireless control units, the laser spot can synchronous drive the 3D simulated-globe by control cursor movement and issued a variety of computer command function. Finally, this work complete and achieve an efficient and economical interactive simulated- globe system.

V. CONCLUSIONS

This work can take advantage of simple technology and classroom existing equipment to employ project the earth or other planets software the images in three-dimensional sphere surface, thereby forming a three-dimensional earth or plants images. By use the interactive technology of this work could make general planet software, used in a variety of panel displays on the projection turn into the three-dimensional imaging sphere. Through the laser spot emitted by the laser emission unit and wireless control unit, allows the laser spot synchronization to control the cursor and software on the three-dimensional imaging sphere for a variety of interactive, and then perform the effect of three-dimensional interactive globe system in any classroom or astronomical museum for formal and social education**.** We have succeeded to develop a useful technique for the classroom and museum, and have obtained the solution of the man-machine interface for its interactive operations.

ACKNOWLEDGMENT

The authors wish to thank the Aim for the Top University (ATU) project of the National Taiwan Normal University (NTNU) for financing this study.

REFERENCES

[1] Albanese, A., Neves, M. C. D., & Vicentini, M. （1997）. Models in science and in education: Acritical review of research on students' ideas about the earth and its place in the universe. *Science and Education*, 6, 573-590.

[2] Crews, W. E. （1990）. *Development of a paper-and-pencil instrument to elicit student conceptsconcerning the earth as a planet*. Paper presented at the Annual Meeting of the NationalAssociation for Research in Science Teaching. （ERIC Document Reproduction Service No. ED324 191）

[3] Vosniadou, S., & Brewer, W. F. （1994b）. Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535-585.

[4] Bednarz, S. W. (2004). Geographic Information Systems: A Tool to Support Geography and Environmental Education?, *GeoJournal*, *60*, 191-199.

[5] Butler, D. (2006). Virtual globes The web-wide world, *Nature Publishing Group*, *439*,776-778.

[6] Demirci, A. (2009). How do Teachers Approach New Technologies Geography Teachers' Attitudes towards Geographic Information Systems (GIS), *European Journal of Educational Studies*, *1*(1), 43-53.

[7] Computer presentation system and method with optical tracking of wireless pointer, U.S. Patent No. 6,275,214.

[8] Non-contact X, Y digitizer using two dynamic ram imagers,U.S. Patent No. 4,507,557.

[9] Three-Dimensional(3D) image projection, U.S. Patent No. 0,256,302.

[10] Three-Dimensional internal back-projection system and method for using the same, U.S. Patent No. 8,066,378.

[11] Display system having a three-dimensional convex display surface ,U.S. Patent No. 7,352,340.

[12] Three-dimensional internal projection system, U.S. Patent No.: 0,027,622.

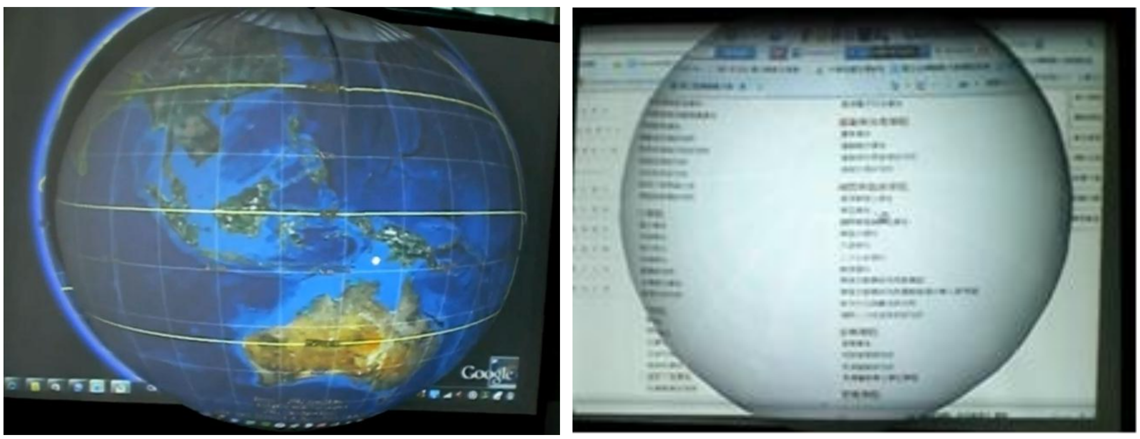


Figure 1. The 3D hemispheric body imaging is to display the output image from data processing unit

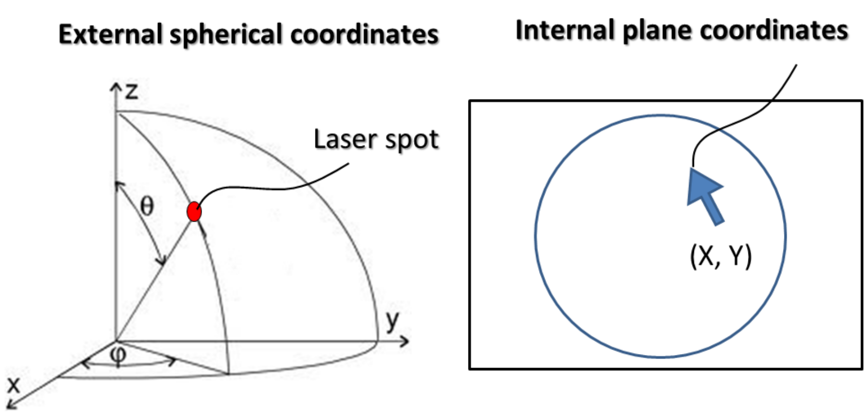


Figure2. Internal coordinates with external coordinate converter operation

 ISBN: 978-1-61208-250-9

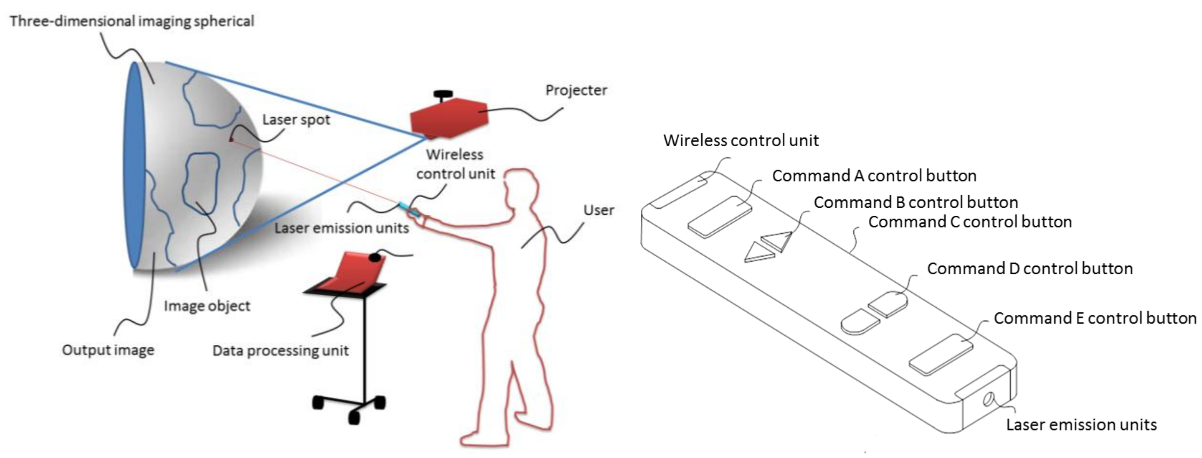


Figure 3. The schematic diagram of 3D interactive simulated-globe system

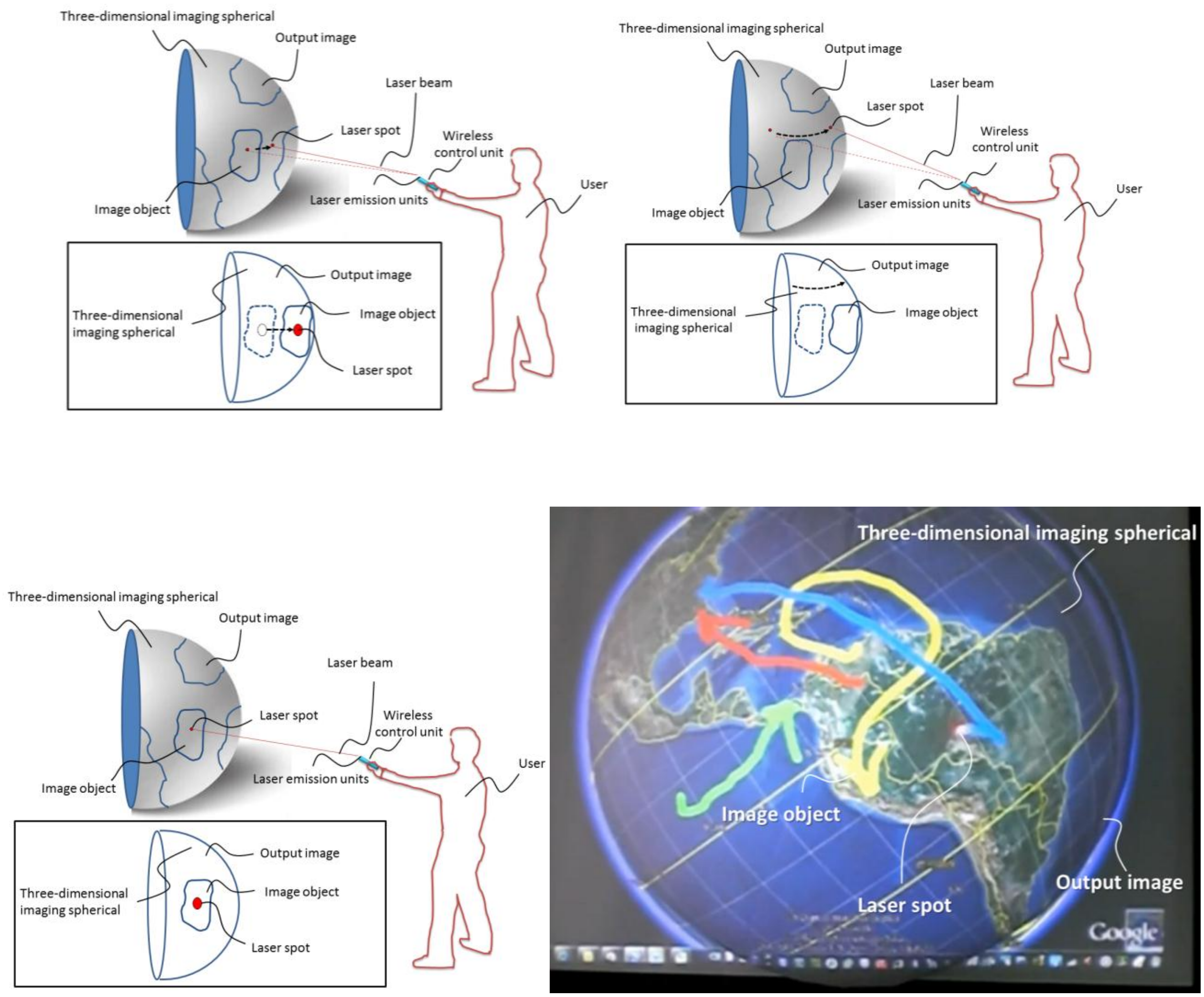


Figure 4. The operator can rotate the earth or the planet, drag, draw , click, move, zoom in and zoom out any icon or function to directly interactive with 3D simulated-globe

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-250-9

Virtual Simulation of the Construction Activity

Bridge Decks Composed of Precast Beams

Luís Viana
Technical University of Lisbon
Dep. of Civil Engineering and Architecture
Lisbon, Portugal
e-mail: luis.viana7@gmail.com

Alcinia Z. Sampaio
Technical University of Lisbon
Dep. of Civil Engineering and Architecture
Lisbon, Portugal
e-mail: zita@civil.ist.utl.pt

***Abstract*— In the execution of bridge or overpass decks several construction processes are applied. A geometric model 4D (3D + time) in a Virtual Reality environment which simulates the construction of a bridge deck composed of precast beams was implemented. The model allows viewing and interaction with the various steps and the main elements involved in the construction process. In order to develop the virtual model, the components of the construction, the steps inherent in the process and its follow-up and the type and mode of operation of the required equipment were initially examined, in detail. Based on this study, the 3D geometric modeling of the different elements that make up the site was created and a schedule that would simulate an interactive mode of construction activity was established. As the model is interactive, it allows the user to have access to different stages of the construction process, thus allowing different views in time and in space throughout the development of construction work, thereby supporting understanding of this constructive method. Since the model is didactic in character it can be used to support the training of students and professionals in the field of Bridge Construction.**

Keywords - Bridge construction; interaction; simulation; virtual reality.

I. INTRODUCTION

In Civil Engineering, there are several construction methods for the execution of bridge decks. This study analyses the constructive method applied to bridge decks using precast beams. In Civil Engineering prefabricated elements are frequently used, because they offer several advantages in urban areas, in works over railway lines, and in general in areas where the placement of trusses is difficult, as this allows quick and economical construction without generating significant local constraints.

Using prefabrication in bridges presents several advantages, such as: the good quality of the concrete of the components produced; economic benefits that result from the use of optimized and standard solutions, which can be used repeatedly throughout the whole process; reduction of congestion on the construction site and the shortening of time-limits for construction, and finally, greater security because it reduces the number of tasks to be carried out on site [1].

This present work aims to contribute to the dissemination of this methodology of construction through a visual simulation created in a virtual environment and to draw attention to its usefulness as a teaching tool, supporting, as it does, the in-depth understanding of this process. For the creation of a visual simulation application Virtual Reality (VR) technology was applied. This technology offers advantages in communication allowing the user to interact with the 4D model as it allows access to different modes of viewing the model in space and time [2].

In order to create models, which could visually simulate the progressive sequence of the process and allowing interaction with it, techniques of virtual reality were used.

When modelling 3D environments there must be a clear idea of what should be shown, since the objects to be displayed and the details of each must be appropriate to the goal the teacher or designer wants to achieve with the model. The use of virtual reality techniques in the development of these didactic applications is, also, generally beneficial to education in that it improves the efficiency of the models in the way it allows interactivity with the simulated environment of each activity. The virtual model can be manipulated interactively allowing the teacher or student to monitor the physical evolution of the work and the construction activities inherent in its progression. This type of model allows the participant to interact in an intuitive manner with 3D space, and to repeat the sequence or task until the desired level of proficiency or skill has been achieved always performing in a safe environment. Therefore, it can be seen that this new concept of VR technology applied to didactic models brings new perspectives to the teaching of subjects in Civil Engineering education [3].

The model discussed here follows other VR models developed within the Technical University of Lisbon at the Department of Civil Engineering, concerning construction works (Figure 1): The models relate to the cantilever process and the incremental launching method of bridge deck construction [4].

II. CONSTRUCTIVE PROCESS OF BRIDGE DECKS

The construction of bridge decks composed of prefabricated beams uses an equidistant distribution of isolated elements placed side by side, complemented by a slab that establishes continuity on the surface of the deck.

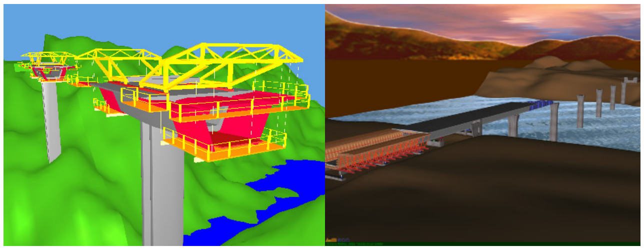


Figure 1. Didactic virtual models of bridge deck construction.

The prefabricated beams are usually built with a length equal to the bridge spans, each consisting of several beams connected from above by a slab concreted “in situ”, and crosswise by transversal beams located on the support. The slab can be made “in situ” using a false-work or pre-slabs, which can contribute to structural strength or assist only as formwork while the slab of the deck is concreted. The most common cross sections in these types of beams are I- shaped (Figure 2) or, sometimes, U-shaped. The shape of the section is determined by various constraints, such as: the procedure of manufacture; the pre-stressing system used (pre- or post-tensioned); transport and assembly; the construction method of the bridge deck slab.

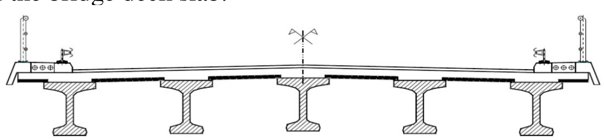


Figure 2. Cross section of a bridge with "I" beams [5].

The constructive method applied to bridges with prefabricated beams can include differences in placement of prefabricated beams, or the type of connection between elements and execution of slabs. The first step consists of placing the prefabricated beams, which can be carried out by means of cranes a launcher (Figure 3).

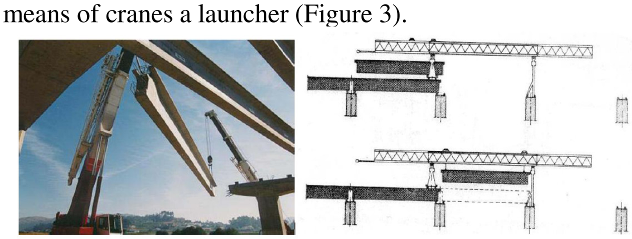


Figure 3. Placing of precast beams on the pillars [6].

The constructive method applied to bridges with prefabricated beams can include differences in placement of prefabricated beams, on the type of connection between elements and implementation of slabs. The first step consists in placing the prefabricated beams, which can be effected by means of cranes or by means of a launcher. After placing the prefabricated beams in its final position, can be used several solutions of connection between these elements and structure: isostatic spans (consist of separate sections, separated by expansion joints, simply supported on the pillars); isostatic spans with solid slab over the joins (prefabricated beams are assembled on supports in an independent way, and a slab in concreted over the joints defining the continuity of the overall deck); hiperstatic solid spans with ordinary reinforcements (continuity is performed by placing the longitudinal slab continuous reinforcements, on the supports, and the space between beams is concreted).

After placing the prefabricated beams in their final position, the connection between these elements is performed using pre-slabs: This method consists of replacing the shuttering and supporting structure of the previous solution for reinforced concrete or pre-stressed slabs with a thickness that usually varies between 6 cm and 10 cm. These pre-slabs can be used as lost shuttering, - in construction they can be used only to support the concrete slab or as bi-functional formwork, that is, functioning as formwork during the constructive phase, but as reinforcement during service.

III. Geometric Modelling and Equipment

After different kinds of bridge deck construction methods had been analyzed, a commonly used model was chosen for the implementation of the 4D (3D + time) virtual model. The deck is composed of beams with I-shaped cross section, lifted by cranes and supplemented with composite pre-slabs. Initially, three-dimensional (3D) geometric modeling of all elements necessary to the implementation of the desired visual simulation was created. The example modeled corresponds to a bridge with the highway profile composed by five spans with the extension of 30m (central) and 24m (lateral) [7]. The cross-section of the bridge deck consists of 4 precast beams (Figure 4).

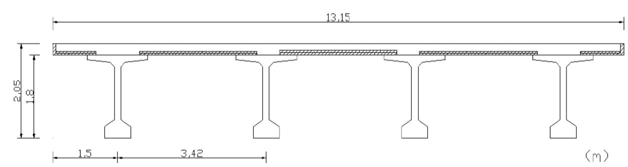


Figure 4. Cross section of the deck.

The 3D modeling process is initiated with the generation of the surroundings of the working place, followed by the pillars, stair towers, worker platforms, provisional and definitive supports and two cranes needed to lift the precast beams.

Figure 5 represents the configuration of the cross-section of the prefabricated beam used and the respective 3D model. In the projection it is possible to see: the running boards (red), protruding out of the beam, which serve to provide resistance and support for the connection of concrete of different ages of the deck slab; the reinforcements connecting precast beams (yellow); other reinforcements needed for lifting the beam (blue).

For the execution of the slab composite pre-slabs were chosen. The dimensions applied in creating the 3D model of the pre-slabs were established based on the drawings of the design of Beira Interior viaducts. Differentiation can be made between pre-slabs, depending on their location on the cross-section: type 1, between beams or type 2, in the console.

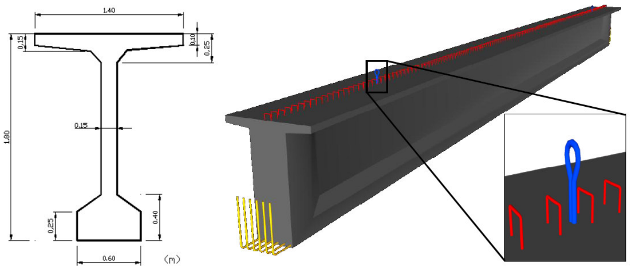


Figure 5. Cross section and prefabricated beam model.

Figure 6 shows a picture of the placement of both types of pre-slabs. The virtual model includes two types of pre-slabs: central and cantilever.

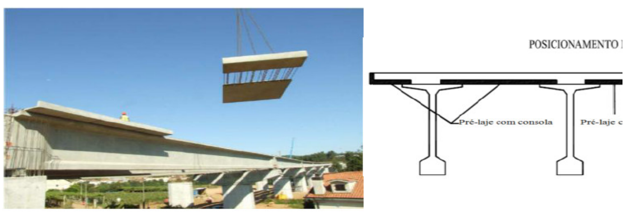


Figure 6. Placement and position of pre-slabs on the deck.

Figure 7 illustrates the geometric model of two kinds of pre-slabs used in the virtual model. The pre-slab in the console is placed in the side of the deck cross-section and the central pre-slab is placed between beams.

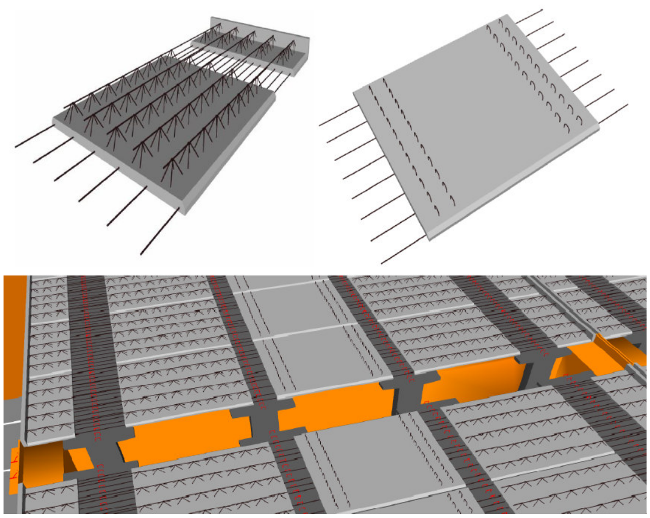


Figure 7. 3D models of the pre-slabs.

IV. THE CONSTRUCTIVE PROCESS ACTIVITY

The visual simulation of the construction was accomplished using software based on the Virtual Reality (VR) technology, the Eon Studio (www.eonreality.com). In accordance with the programmed sequence of the construction:

- The virtual simulation starts with the presentation of the work place environment;
- The simulation of the construction activity continues with the insertion of additional elements, such as the stair towers (for access to the top of the pillars) and the work platforms (which allow the workers to move around and complete their tasks) (Figure 8).

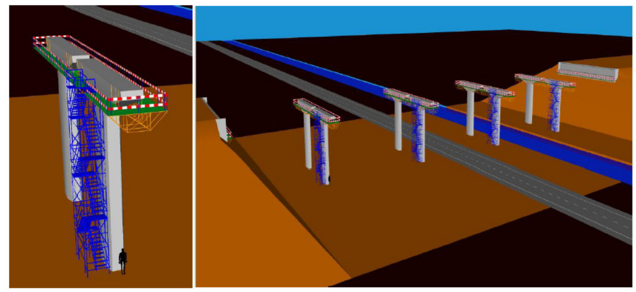


Figure 8. Display stair towers and work platforms.

- In the virtual space the placing of the definitive support device is simulated, followed by the placement of the temporary support, on the top of the pillars (Figure 9);

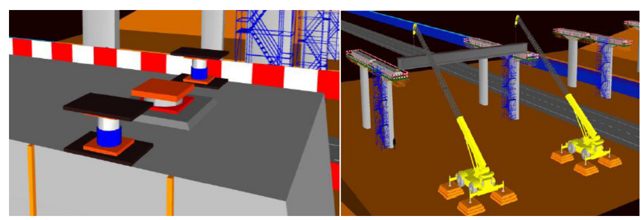


Figure 9. Placing of definitive and provisional support and the prefabricated beam.

- Each beam is raised by two cranes and placed on the temporary support devices (Figure 9);
- The simulation of construction process proceeds with the placement of the pre-slabs on the prefabricated beams (Figure 10);

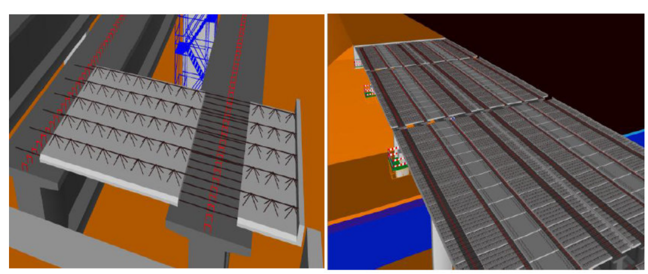


Figure 10. Placement of pre-slabs and concrete transversal beam and slab.

- At this stage, it was found that due to the large amount of reinforcements set in the pre-slabs (Figure 11), the movement of the camera (from the point of view of the user) through the virtual environment was very slow because the drawing file was already too heavy. It was, therefore, decided not to display 3D model of the pre-slab reinforcements and also the brackets from prefabricated beams;

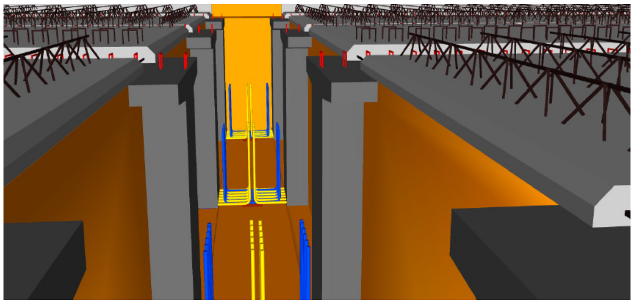


Figure 11. Large amount of reinforces set in the pre-slabs.

- Then, the reinforcement of the slab is placed over the pre slabs and the deck slab is concreted (Figure 12);

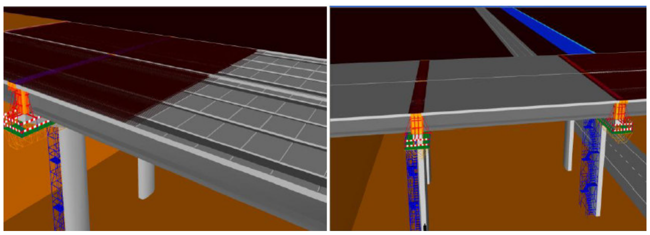


Figure 12. Placing reinforcement and concreting the deck slab.

- Next, the transversal beams are concreted. Figure 13 illustrates the reinforcements, the 3D model and the placement of the formwork and reinforcement of one of the transversal beam;

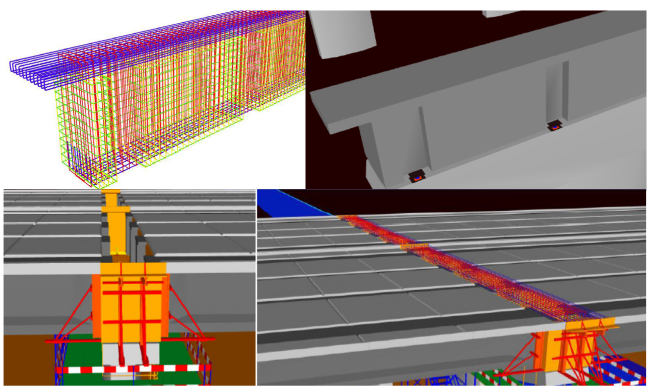


Figure 13. Concreting one of the transversal beam.

- After the completion of the construction of the deck the provisional support devices are removed (Figure 14);

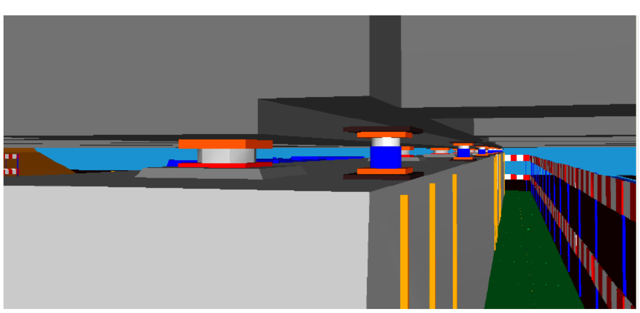


Figure 14. Removing the provisional support devices.

- Finally all complementary elements necessary for the road traffic were inserted into the top of the deck (Figure 15).

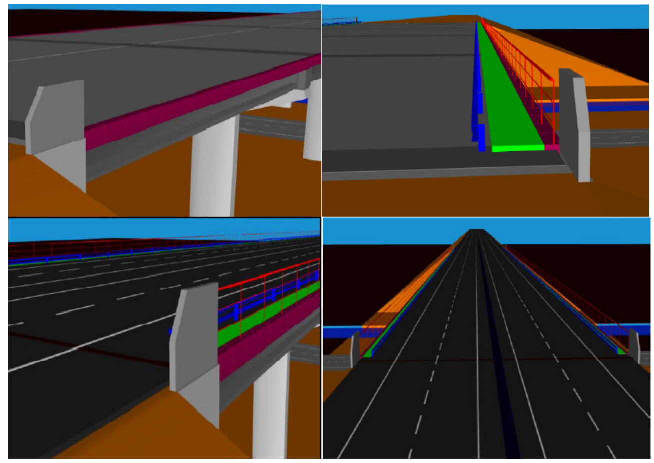


Figure 15. Placing the complementary elements.

- The complete bridge can now be observed from any point of view (Figure 16). The model allows the user to use the zoom sufficiently well in order to understand the final configuration of the bridge.

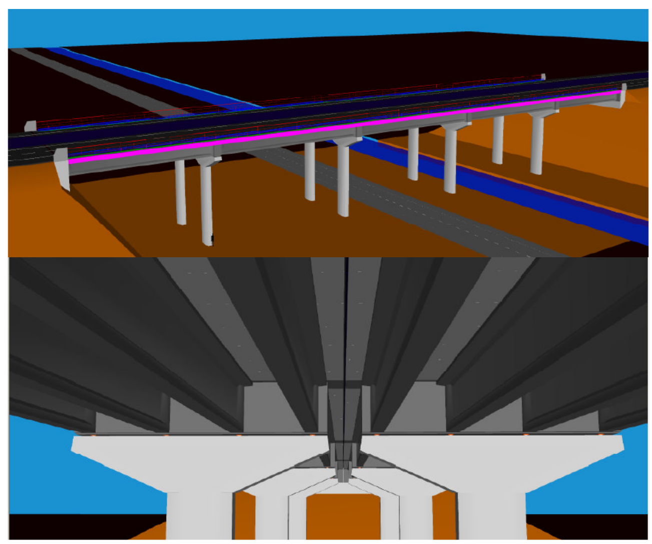


Figure 16. Views of the complet deck.

V. Educational Aspects

The virtual model is currently used in the discipline of Bridges. The introduction of the topic on the construction processes used in bridges, in particular, the methods related to bridges composed of precast beams is supported with the visual and animated simulation of the modeled construction deck. The teacher of bridges, which refers to the use of the virtual classroom model, verifies that the student is quite observant with the lively presentation of the entire process of construction of the bridge. All constructive and the sequence type and geometric form of constructive elements required in the execution of the deck are presented in the virtual model with enough clarity.

The model was implemented with great accuracy in terms of scientific content in order to obtain a correct model technically and so is effective as a teaching model. The target student of the virtual application is the student of the last year of the course of Civil Engineering, in particular of the area of Structures, as is the branch that is more related to the general theme of bridges. The constructive process of bridges is a theme that is fundamental in the teaching of design and types of bridges. This model and other previously performed (incremental method and method of advances [4]) composes together with this virtual model, a set of educational tools, based on IT (information technology) and VR (virtual reality), which serve as support for the discipline of Bridges. The models were made available to students in the www sites of Bridges and research related to masters' theses. There request and manipulation are made available to any student.

VI. CONCLUSIONS

This paper analyzes some constructive processes concerning bridge deck formed of precast beams and describes the implementation of an interactive model that simulates the construction work activity. The virtual application shows one of the methods most often applied in the construction of this type of Deck Bridge. In the creation of the model software based on the RV technology was used.

Virtual Reality allows, through interaction with 3D models of the environment, representing building components and equipment, the creation of the constructive sequence in time and space by simulating the progression of the construction of the deck, which allows a good understanding of the whole process.

The model 4D (3D + time) offers several advantages, allowing a deeper awareness of the relationship of the components of the building and the phasing of the work, leading to a better understanding of the spatial movement of equipment and of the component placement in work. Since the traditional designs of graphic documentation of the construction project are sometimes more difficult to understand, it can be seen that this model is clearly didactic in character and as such can be used to support the training of students and professionals in the field of Bridges.

REFERENCES

[1] J. N. Câmara, "Prefabrication of Bridges and Viaducts", National Meeting of Prefabricated Structures, Prefabricated Company Castelo, Lisbon, Portugal, January 18, 2001.

[2] S. Woksepp, "Virtual Reality in Construction. Tools, Methods and Processes", PhD thesis in Civil Engineering, Luleå University of Technology, Sweden, November 2007.

[3] A. Z. Sampaio and P. G. Henriques, "Virtual Reality Technology Applied to the Visual Simulation of Construction Activities",The Open Construction & Building Technology Journal, Mathew Honan, Editorial Director, Bentham Science Publishers Ltd., vol. 2, pp. 7-14, 2008.

[4] A. Z. Sampaio, P. G. Henriques, P. Studer and R. Luizi, "Interaction with Virtual 3D Models in Order to Access the Construction Processes of a Wall and a Bridge", CC-2005, The 10th Int. Conf. on Civil, Structural and Environmental Engineering Computing, ed. B. Topping, Rome, Italy, 2005.

[5] C. F. F. Sousa, "Structural Continuity in bridge decks built with Prefabricated Beams. Solutions with Reinforced Concrete Connection". MSc thesis in Civil Engineering Structures, FEUP, Porto, Portugal2004.

[6] F. Leonhardt, "Basic Principles of the Construction of Concrete Bridges", Volume 6, Interciencia, Rio Janeiro, Brasil, 1982.

[7] A. J. Reis, "Bridges", Didactic text, ed. AEIST, Lisbon, Portugal, 2006.

The iPad in a Classroom: A Cool Personal Item or Simply an Educational Tool?

Andrea A. Gasparini
Digital Services, University of Oslo Library
University of Oslo
Oslo, Norway
e-mail: a.a.gasparini@ub.uio.no

Alma L. Culén
Department of Informatics
University of Oslo
Oslo, Norway
e-mail: almira@ifi.uio.no

***Abstract* — In this paper, we discuss the dual role of the iPad among the teenage high school students using the tablet as a 1-1 (one tablet per student) educational tool. On one hand, the iPad is a personal, mobile and cool piece of technology. On the other hand, it is a piece of technology provided by the school and given to students as anytime, anyplace mobile educational tool. Our goal is to understand the space between the use situations related to school work and those that are private and personal. After ten months of observations of the use of the iPad, we conclude with that the iPad is treated by teenage participants in our study as purely educational tool.**

Keywords - cool; identity; iPad; education; learning; techno-cools.

I. INTRODUCTION

When the iPad was launched in the 2010, numerous iPad oriented projects and studies emerged worldwide within diversity of settings. Specifically, for the field of education, the interest was, according to Apple [1], reaching unprecedented heights.

This new artifact came quickly to represent a promise of a paradigm change in learning, nearly a promise of a revolution in the field of education [5]. It also represented a possibility to change a known trend of classroom technology non-acceptance [8, 20] and become a tool that proves that technology can bring considerable benefits to education [1].

Creativity was expected to play a major role in moving the education towards more constructivist practices [7]. The iPad has a large number of applications supporting productivity and creativity. In using them, students may open up a possibility to take a more active part in their own education, by, for example, designing a part of their own curriculum through the use of different apps.

However, creativity in the digital era is quite complex. It involves navigating in a plethora of platforms, channels and applications, all of which may demand some learning and mastery. A cognitive aspect of multimodal representations of content in the learning processes typically supported on the iPad [4] requires a new kind of literacy to deal with, both for teachers and students. The multimodality theory advocates an understanding that communication, when using a device such as the iPad, occurs through multiple but synchronous modes such as images and graphics from the camera or the Internet; touch using the touch screen, and audio input and output. This problematic is not new. For instance, Sneller describes an earlier study [25] done with a class where students used a tablet PC. The study showed that students had a positive attitude towards the use of the tool that supported their active learning styles. It also showed that much more work was required from the teacher; the effective use of the tool implied a major redesign of the curriculum to be taught in a classroom. Sneller concludes with *"Even on those days when you feel like this challenge is akin to that involved in herding cats, the victories will come."*

An obvious use of the iPad for students is to use it as an e-book reader and read books on this electronic platform. An increasing part of new e-books, including the academic ones, developed for the iPad or other electronic reading devices offer brand new features. Some examples of those features include direct interaction with the content, word search, video or images access, hyperlinks to references etc. This kind of reading can be described as is an intermittent, digressive and collective act [14], where, in spatial transitions, remembering words becomes very difficult to achieve or is no longer possible [14]. For students, the book is transitioning from a product to a service.

Given that the cognitive complexity has increased and different kinds of skills are needed to deal with new, multimodal, mobile learning tools, it is difficult to say if the iPad actually enables increased creativity and production and in which ways it does so. There are three angles of approach to investigation of this: studying the problem with the tool in focus, focusing on users involved in a use situation and a combination of both. The later is the approach we chose in the study described in this paper.

In our previous study with the iPad as an educational tool for geosciences university students [11], we found out that productivity and creativity were strongly related to various ownership issues such as time investment into mastery of the device one does not own, cost of applications that are to be installed on the device that one does not own, proprietary software that is hard to modify and customize, issues around annotating curriculum related articles, ownership/copy rights issues related to that, etc.. Additionally, factors from the educational Technology Acceptance Model (TAM) [27] such as perceived usefulness, perceived ease of use and teachers influence were found also relevant for the use of the iPad in the classroom ecology.

The study described in this paper takes into account users, the use situation and the tool. The users are technology savvy students, also referred to as the Net generation, digital natives, millennials, etc. These ever more mobile students

 ISBN: 978-1-61208-250-9

actively use different technologies offering them content access anyplace, anytime. The technologies such as smart phones and tablets have become a new arena for the "cool" products, making the Net generation students more aware of the choices they make. Coolness is a factor that this youth responds to [9, 19, 21]. As an attribute of the product, it is somewhat elusive and paradoxical. In [9], the term techno-cool is used to describe such products. Clearly, designing techno-cools is different from designing cool products that are not technology based. The design and production of technology must strike a balance between short-term, fast changing product versus long-term product and brand building. Most large, mature technology producers trying to cater to the "cool" market understand the need for branding efforts that give sustainable, long-term advantage. Thus, really well designed products may become what we call mainstream techno-cools, the iPad representing one of them. The high number of sold iPads can not only be understood as a demand from the market for this device, the coolness effect was also a factor [2]. In their turn, researchers in interaction design have recently begun to study coolness as a factor to take into account when designing new technology [9, 16, 24]. It is interesting that Holtzblatt [16] proposes a model for designing cool innovative products with joy, identity, connectivity, sensation and accomplishment as drivers, coupled with use situations (see Fig. 1). The components in a wheel of joy are precisely those that are generally important for the Net generation we are considering in our study.

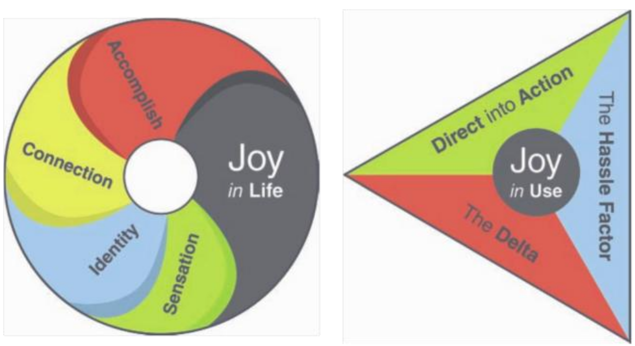


Figure 1. Holtzblatt's wheel of joy and a triangle of design.

The group perception of coolness is obviously important in the social context for these young people at school, work, home, or out and about with friends. In [9] we define some attributes of coolness relevant for the iPad:

"*However, if one invests in a piece of technology, one expects it to do what it is designed for and much more, to be useful and easy to use, almost never unique, but possibly customizable and definitely, one expects it to be fun.*" Attributes making a product cool like usefulness, ease of use, mainstream techno-cools, connectivity, identity, fun and happiness are thus representing the appropriate context investigate in relation to "cool" educational technologies we well. Coolness can also be related to different attitudes young people may have, from rebellious [23] to more laid back [26].

Related to the use of the iPad and identity, it is of importance how users personalize this artifact. More and more of our lives is now kept and represented by a large group of software in the cloud. Users need then to take a stand where, when and how they shall be in touch with this "second", "cloud" representation of their lives. Turkle [28] advocates and defines this representation as the "second self". In the earlier 1980s work, Turkle [28] defined how then new, personal computer affected the identity of the user and her angle of studying it:

"…my focus here is on something different, on the "subjective computer". This is the machine as it enters into social life and psychological development, the computer as it affects the way we think, and especially the way we think about ourselves."

The iPad has the possibility to acquire and reflect a number of attributes, and represent the content of our second self in a new wrapping, re-sorted, re-made, sometimes better than the original, almost as a simulacrum [3]. For the user, the iPad can be a new companion, with all the positive and negative aspects this companionship involves.

In this paper, we report from a case study in a high school, where we focus on the personal and the educational use of the iPad. Factors such as self identity and coolness affect these uses. We also recognize the importance of factors like joy and fun as mandatory for a positive user experience with the device. Yet, this study shows that students prefer to use the device mostly for educational purposes.

The paper is organized in five sections. In the next section, Section 2, we describe our case study. In Section 3, we present our methods and findings, followed by Section 4, where we use an identity and coolness driven approach to discuss our results. Section 5 concludes the paper.

II. The Case Study

Our case study involved teenagers, aged 16-17 and their use of the iPad both as a personal mobile device, and as a mobile educational tool. They are the generation that has grown up with technology and are being educated with it [10, 17, 22]. This research project was a joint venture between different actors: a private school, a representative from Apple, a company offering help to introduce technology in education, and us as researchers. The whole project had a broader purpose of seeing if the iPad could help students in their studies, if it is a game changer in education [5]. For this case study, as mentioned in the introduction, we focused on relation between the identity (including personalization) and coolness (including fun, joy, cool behaviors, etc.) and how they influence private vs. educational use. Since in Norway every high school student receives both school books and a laptop for free during their studies, the project also gave us the possibility to explore the following question: is the iPad cooler and more personal than the laptop?

The private high school at which the study took place is in the situated very central, in the heart of a major city in Norway. A class of 25 first year students was chosen, and after two preliminary stakeholders meetings in January of

2012, we started our study. A study was of one-student-per-iPad type. The students have received iPads 2, preloaded with about 200 applications, as well as an external keyboard and a small cover.

At the start of the spring semester (January – June, 2012), all the students had a week long support from the company facilitating the introduction of the iPad in education (see Fig. 2). During this week, it was explained how the iPad and all of the preloaded apps worked. The support from this company did not stop there. The school organized monthly visit by the same team for the rest of the semester.

The classroom was also equipped with an Apple TV system in order to provide students with the possibility to present their work for the whole class directly from their iPads. The classroom, and the rest of the school, had a wireless connection to the Internet.

The teachers had received the iPad four weeks before the students, at the end of the fall semester and had time during the Christmas break to play with the device, explore it both as a personal item and as a tool it was soon going to be. They too received some intensive course training at the start of the spring semester. The focus for teachers was on learning how to use different apps for re-creating the curriculum, when possible, and how to use them in class teaching. For instance, iMovie was recommended as a tool for recording homework assignments and iThoughts for making overviews of diverse subjects and assignments.

The curriculum books were delivered by the publishing houses in PDF format, and made available to the students to download. For the school this was a large logistic gain. Moreover, for the IT support center of the school the migration from laptops to iPads was a definitive improvement since the iPads were much easier to support and manage. It was planned to deploy one server as a backup of students' iPads. At the end of the semester this server was still not online, illustrating the fact that standard issues with technology, including non-deployment and break-downs were part of this study as well.

As far as we know, there were no restrictions or instructions given to either students or teachers regarding the use of the iPads outside the educational framework.

III. METHODOLOGY AND FINDINGS

A total of five full school day observations were carried out. In addition, a series of interviews were conducted. Nearly all students and faculty were interviewed both at the beginning and at the end of the semester. With one especially interested student, whom we will call Stella, observations and interviews with the family members were carried out in the home environment.

The visits to the school were divided in two periods, one at the start of the semester (3 days) and one at the end (2 days). In addition to observations and interviews, two surveys were carried out. The first survey explored students' perception of the iPad as a techno-cool artifact and contained questions around private use as well. The second survey, carried out at the end of the semester, covered a range of subjects, from the use of the curriculum on the iPad to the Learning Management System (LMS) of the school, i.e., focused on the use of the iPad as an educational tool.

At the start of the study, almost all teachers were positive toward introduction of the iPad into classroom ecology and had envisioned new ways of teaching using the iPad in the classroom. Having all the books available at all times in class, organization and systematization of their notes, making of subject mind-maps, and extensive use of movies were all new and exciting. One teacher said:

"*If they use pen and paper for taking notes, the next time they have a class they have either forgotten what the notes were about or they forgot to take the notes to school with them. Now, using iThought, it is easier to check if they have done it right and then it is certain that they have it ready for the next session*" (author's translation). The euphoric mood around the iPad and remarks on its coolness were also present among the teachers. One of them stated in an interview: "*I like my iPhone, but I love my iPad!*"

Figure 2. Introductory workshop for students on use of the iPad.

Introducing new technology in the classroom ecology was also difficult for the teachers. New plans for the curriculum required new expertise. The use of camera, touch screen, and microphone in various ways by various applications represented a challenge. One of the teachers made an interesting remark related to how to prepare the students to deliver homework. If the teacher chooses an app, then it is mandatory for him/her to test if it is possible to use the app correctly towards specific learning goal, and prepare some potential tips to give them. Another major issue was related to whether students will be able to, or not, learn the curriculum using the iPad. The aforementioned multimodal literacy, bringing a new pedagogy into the classroom, requires a huge effort, thus confirming the findings of [25].

The end of the semester interview period with teachers still showed a positive attitude. What seemed to be appealing to the teachers was the possibility the iPad had to provide a greater variety in the way students can deliver their work. Allowing the students to use videos, drawings, comic strips and mind maps led a larger number of students to take a more active part in the classroom life. This view was also confirmed by observations: the students were indeed more active in class at the end of the semester than they were at the beginning. The possibility of presenting the results of their work, sharing the results and observing what others have made, invited all the participants to repeat and have a greater focus on the content of the lesson. Interviews with the students at the start of the project revealed a very enthusiastic approach to this new technology. Some the early feedback from students also touched upon well known problems [13] such as the lack of Flash support. In particular, the students found a lot of appropriate content on the web for science classes which was not available on the iPad. Using the iPad for mathematics homework proved to be difficult also, since it was not so easy to write fast all the calculations. One student, Stella, found a workaround: taking pictures with the iPad and inserting them in the homework (see Fig. 3) using Notability.

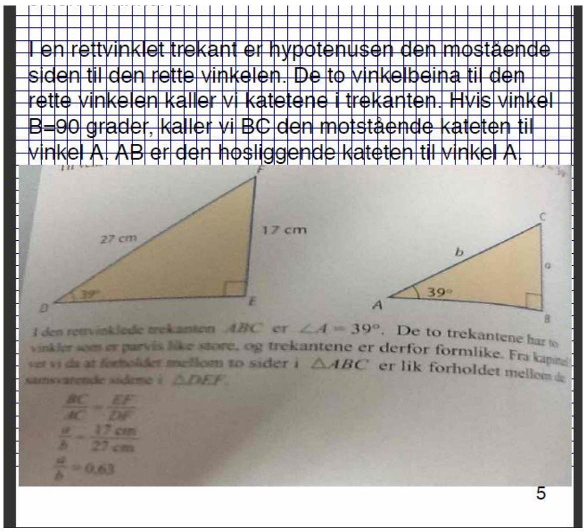


Figure 3. New ways of doing homework in mathematics

From the very start Stella was unique; she loved the iPad, and she loved using it. She gives the iPad credit for improving her grades and a feeling a larger degree of mastery over her school days. She used the iThought to organize all the content for all her different subjects. For instance, Fig. 4 shows an excerpt of a map where 200 keywords for a specific subject are plotted. Stella had some problems with her vision, so reading text was sometimes difficult. Enlarging fonts gave the possibility to read with much less strain on her eyes. She also duplicated, to the largest extent possible, the apps she had on her iPad on her iPhone, thus making it possible to do homework truly any-place, any-time. Stella was so engaged with the iPad, the principal of the school asked her to make an introductory video presentation of the iPad for education that was to be shown to the next generation of students. At the end of the semester she was still using her iPad for the school work, but a change occurred. When it came to leisure use, the PC was still necessary to satisfy her needs. To update the iPhone with iTunes or to see movies, the iPad did not function optimally. She tried some workarounds with mixed results. The end result was the use of the iPad for education and of the PC for leisure. She told us that this was actually her impression for the rest of the class too:

"*I got the impression that most of the students do not use the iPad for leisure. This is particularly true for those that play games a lot.*"

During other interviews at the end of the semester the rest of the students confirmed Stella's statement; the use of the iPad was marginalized to school work only. As reasons for this "go back" effect, students had problem to explain what motivated them to do so. One girl, deeply reflecting over the issue, managed to say the following:

"*Maybe this is so for me because I am old fashioned. For me, the laptop is better than the iPad. At the start, the iPad was really cool, but now, it is so related to school work and learning. Thus, I prefer the laptop since I do not use it at school. When I use it, I know that it is for fun*".

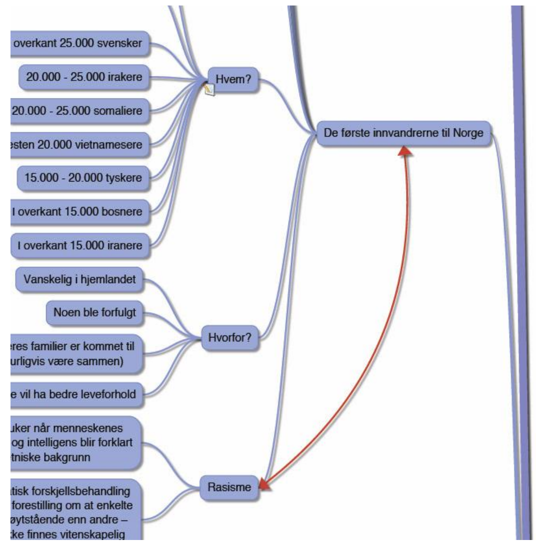


Figure 4. A portion of iThought map for Stella's social science class.

The results of the end of the semester survey supported this finding: only 1 out of 10 respondents said that he/she uses the iPad more than needed for school work, 2 did not know what the distribution of hours between school and home use was. The remaining respondents said that their school use equals their total use, Fig. 5.

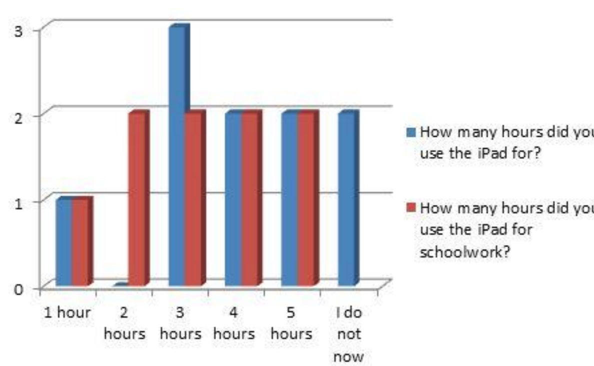


Figure 5. Total time of iPad use per day vs. the time for school work.

IV. Discussion

The most interesting finding from this case study was the emergence of the need for separating the educational and leisure use of the iPad. Students had some issues against the use of the same technology in the educational and in the leisure contexts. We have observed the same phenomena in our elementary school study [10, 13] with 5th and 6th grade students. The effect though, was a reverse one: they all used the iPads for play and leisure, while at school, the iPads ended up on the classroom shelves, unused. The reasons for this reversal may be many, for us, it suffices to note that the separation of use spheres occurred in both studies.

Other studies [6, 15, 18] have also revealed a possible resistance students may have against personalization and appropriation of technology that supports their learning. Hossein [15] also points out a possible reason for this defiance students may have against the personalization of learning technologies: "*Younger students may see this as an encroachment into their recreation and resent the educational provider for taking over their space unless the technology can seamlessly be integrated and be seen by the students as a hybrid of a recreational and study tool*." Our findings support this understanding. For instance Stella says in one of the interviews: *"It* (iPad) *has applications for everything your PC has. Well, almost everything. At least, it is certainly enough to work with it effectively at school, and that is what it should mainly be used for."*

Next, we consider the identity and reflexivity issues and their influence on this split. This discussion is followed by the discussion of the influence of coolness.

A. Indetity and reflexivity

Our findings support the need to define the identity of the users in the context of new technologies, both in an educational context and in the private sphere.

We have mentioned the "second self" as a possible approach to understanding the dynamics in the behavior of the students. From Turkle's '80 representation until today, a major change has occurred; our identity is more and more placed in the cloud. Facebook, Twitter and other services are indirectly a representation of ourselves, our *second self*, and the iPad is one of many windows into that world. Turkle made also a new edition of the book twenty years later [28] and she adds an important observation:

"*In 2004 the cultural message of digital technology is not about simplicity but complexity, not about transparency but opacity*".

This description is representative for the problems users may have to deal with, when relating to this second self. Complexity and opacity can be factors the iPad amplify using a world of apps that filter and represent the content in a new wrapping, re-sorted and re-made. The user then must take a stand to choose if this new reality is better than the original.

The iPad can act as a projective medium, requiring the user to accept the reflexive effect and deal with this new changed representation of identity. The students in our study selected the option of putting aside the iPad in their spare time. We may understand this behavior as an act of avoidance to deal with this "second self" on the iPad.

It is interesting to note that, connectivity, a need of this Net generation to be online with others present at all times, was not satisfied through schools new social network, called Connect. Connect functions much like the Facebook, except that all interaction is stored on a local server. However, Connect could not replace the Facebook, and was used, again, only to discuss educational issues, not for connecting with peers. The "second selves", the cloud identities of our students "preferred" PCs and the "original" social media.

B. Coolness

Coolness alone was not enough to make the iPad indispensable in both educational and private use. Even when the important attributes such as usefulness, fun and joy in use were all present.

Even though the iPad definitely had a "cool" status according to the data we collected through interviews and the first survey, the iPad did not reach the "I can't go back" point [16]. While the students could not even imagine going back from the iPod to the "Discman" or the "Walkman", they had no problems going back from the iPad to the laptop for their entertainment and "connectivity" time. Moreover, this going back to the laptop happened in a short period of time. Why is the iPad then still considered to be cool?

One could argue from our data that usefulness, fun and enjoyment in use were more dominant in educational, that the iPad was a cool tool to use at school. We further assume that the coolness from one context of use (the school) is then transferred to the device itself (and thus the device is still considered as cool by a large majority of students).

V. Conclusion and Future Work

In the classroom, the iPad came to be viewed by the students as a tool for self-improvement in the educational arena. At the start of the study, the iPad was also used for private purposes, connectivity and self-representation in social media, but this route was quickly abandoned. It seems that students have a need to separate the educational and the personal use of the iPad.

On the other hand, the educational goal of the private school was reached; the migration from the laptop to the iPad was successful and many positive effects in the learning arena were experienced. A period of time will be necessary for the teachers and schools to adapt the curriculum to these new ways of using the device involving images, text and audio. The potential is, according to both teachers and students, large. Future research should focus on deeper and broader investigation of reasons for separating the personal and the educational use of tablets, as well as its implications.

References

[1] Apple. An Apple product for every student. Retrieved January 20, 2013, from http://www.apple.com/education/labs/.

[2] R. Anderson, Social Diffusion of iPhone Technology. Retrieved January 20, 2013, from http://thesocietypages.org /eye/2008/07/.

[3] J. Baudrillard, Symbolic exchange and death. London: Sage, 1993.

[4] N. O. Bernsen. Multimodality theory. In Dimitrios Tzovaras, editor, Multimodal User Interfaces, Signals and Communication Technology, pp. 5–29. Springer Berlin Heidelberg, 2008.

[5] G. Brown-Martin, 2010. Game Changer: Is it iPad?. Retrieved January 20, 2013 from: http://www.handheldlearning.co.uk /content/view/64.

[6] L. Corrin, L. Lockyer, and S. Bennett, "Technological diversity: an investigation of students' technology use in everyday life and academic study," Learning, Media and Technology, 2010, vol. 35(4), pp.387-401.

[7] M. Csikszentmihalyi and R. Wolfe, "New conceptions and research approaches to creativity: Implications of a systems perspective for creativity in education," *International handbook of giftedness and talent*, vol. 2, pp. 81–91, 2000.

[8] L. Cuban, Oversold and Underused: Computers in the Classroom. Harvard University Press, 2003.

[9] A. L. Culén and A. Gasparini, "Situated Techno-Cools: factors that contribute to making technology cool in a given context of use," PsychNology Journal, 2012, vol.10, 2, pp. 117-139. Retrieved Jan 20, 2013 from www.psychnology.org.

[10] A. L. Culén and Gasparini, A. A. "When is a Student-Centered, Technology Supported Learning a Sucess? *International Journal of Digital Information and Wireless Communications (IJDIWC) 2*, 3, 2012, pp. 256–269.

[11] A. L. Culén, B. K. Engen, A. Gasparini and J. Herstad, "The use of iPad in Academic Setting: Ownership Issues in Relation to Technology (Non)Adoption," Old Meets New: Media in Education: Proceedings of the 61st International Council for Educational Media and the XIII International Symposium on Computers in Education, 2011, pp. 555-563.

[12] A. L. Culén and A. Gasparini, "iPad: A New Classroom Technology? A Report From Two Pilot Studies", Information Sciences and e-Society, 2011, pp. 199 – 208.

[13] A. Gasparini and A. L. Culén, "Acceptance factors: an iPad in Classroom Ecology", The International Conference on E-Learning and E-Technologies in Education (ICEEE), September 2012, Lodz, Poland.

[14] M. Gauchet, "Le Livre, le numeriquè," Le débat, 2012, vol. 170, pp. 1-192.

[15] A. Hosein, R. Ramanau, and C. Jones, "Learning and living technologies: a longitudinal study of first year students' frequency and competence in the use of ICT," Learning, Media and Technology, 2010, 35, 4, pp.403-418.

[16] K. Holtzblatt, "What makes things cool?: intentional design for innovation," Interactions, 2011, vol. 18, 6, pp. 40-47.

[17] C. Jones, Networked Learning, Stepping Beyond the Net Generation and Digital Natives. In Exploring the Theory, Pedagogy and Practice of Networked Learning. In L. Dirckinck-Holmfeld, V. Hodgson & D. McConnell (Eds.), (pp. 27-41): Springer New York. 2012.

[18] G. E. Kennedy, "First year students' experiences with technology: Are they really digital natives," Australasian Journal of Educational Technology, 2008, vol.24, 1, pp.108.

[19] C. Nancarrow, P. Nancarrow, J. Page, "An analysis of the concept of cool and its marketing implications," Journal of Consumer Behaviour, 2002, vol.1, 4, pp. 311-322.

[20] C. Norris, T. Sullivan, J. Poirot, E. and Soloway, "No access, no use, no impact: Snapshot surveys of educational technology in K-12, " Journal of Research on Technology in Education, 2003, vol.36, 1, pp. 15-28.

[21] K. A. O'Donnell and D. L. Wardlow, "A theory on the origins of coolness. Advances in Consumer Research," 2010, vol.27,1, pp. 5.

[22] D. G. Oblinger, and J. L. Oblinger, Educating the net generation. Boulder, Colo.: EDUCAUSE, 2005.

[23] D. Pountain, and D. Robins, Cool rules : anatomy of an attitude. London: Reaktion Books. 2000.

[24] J. Read, D. Fitton, B. Cowan, R. Beale, Y., Guo, and M. Horton, "Understanding and designing cool technologies for teenagers," Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, ACM, 2011, pp. 1567-1572.

[25] J. Sneller, "The Tablet PC classroom: Erasing borders, stimulating activity, enhancing communication, " Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007.

[26] P.N. Stearns, American cool : constructing a twentieth century emotional style. New York, NY [u.a.]: New York Univ. Press.1994.

[27] T. Teo, Technology Acceptance in Education: Research and Issues. Sense Publishers, 2001.

[28] S. Turkle, The Second Self: Computers and the Human Spirit. Simon & Schuster, 1984.

Architecture of an Intelligent Tutoring System Applied to the Breast Cancer Based on Ontology, Artificial Neural Networks and Expert Systems

Henrique P. Maffon, Jairo S. Melo, Tamara A. Morais, Patrycia B. Klavdianos, Lourdes M. Brasil
Post Graduation Program in Biomedical Engineering
University of Brasilia at Gama
Brasilia, Brazil
{henrique.maffon, jairossmunb, tamaramorais89,patryciaklav, lmbrasil, }@gmail.com

Thiago L. Amaral
Graduation Course in Engineering
University of Brasilia at Gama
Brasilia, Brazil
thiaguimamaral@gmail.com

Gloria Millaray J. Curilem Saldias
Eletrical Engineering Department
Universidad de La Frontera
Temuco, Chile
millaray@ufro.cl

***Abstract*— This paper presents an Intelligent Tutoring System (ITS) applied to the teaching of breast anatomy and pathology, more specifically of the breast cancer, the type of cancer that kills more women in the world. This paper aims to elucidate the importance of using these systems, lift requirements for the development and designing of an architecture of ITS. Through resources and applications of Artificial Intelligence techniques this ITS has the capacity to acquire the profile of its user and define teaching methodologies to build interactive and dynamics environments, based including in the use of Virtual Reality. The architecture of this ITS consist on four modules: Tutor Module (Artificial Neural Network), Student Module (Expert System), Domain Module (Ontology) and Interface Module (Adaptive Hypermedia Systems). The use of this ITS provides a didactic help to students and health professionals to the understanding of the explanations and practical applications needed to this domain of knowledge.**

Keywords - Intelligent Tutoring System; Expert System; Artificial Neural Network; Ontology.

I. INTRODUCTION

This article presents an Intelligent Tutoring System (ITS) – a system which incorporates Artificial Intelligence (AI) techniques – applied to the breast cancer [1]. This is the most common type of cancer that affects women and causes more deaths among them in the world. In 2008, 23% (1.38 million) of new cases and 14% (458.400) of deaths were caused by this disease [2]. In 2012/ 2013, 52.600 new cases of breast cancer are expected for Brazil, with an estimated risk of 52 cases in each 100 thousand women [3].

In order to provide improvements in both education and quality of life, the areas of health, education and information technology have been working together to develop education systems and disseminate information about prevention of diseases [4].

The first educational software used were the Computer Aided Instruction (CAI) Systems, designed in the 50's with the aim to provide to the students improvements on learning. However, these systems could not differentiate the individual abilities of students, because they used to implement same actions to all users [5].

Carbonell (1970) proposed the creation of the ITS in order to create an environment that allows a more individualized interaction between the student and the system [1], which has proven the effectiveness in improving the performance and motivation of students [6].

The use of AI techniques gives ITS a great capacity to store knowledge about the learning virtual environments and about the apprentice's characteristics. These characteristics are used, on the scope of Education, to define which strategies and tactics can be used to get the best teaching process [7]. The strategies describe the required process to achieve learning and they are based in theories as Behaviorism and Constructivism, among others; the tactics refer to specific resources used to facilitate the teaching, as images, videos, texts, among others features.

Most of develop ITS is composed by four modules (Student, Domain, Tutor and Interface), however this pattern is not always used. Basically, the Student Module stores the apprentice's characteristics; the Domain Module stores the content to be taught; the Tutor Module decides which ways the content shall be presented, adapting the Interface to the apprentice's characteristics, to display the content through an immersive, interactive and intuitive environment [8].

The aim of this article is to elucidate the importance and lift requirements for the development and designing of an architecture of ITS, including its possibility to use this ITS in a Medical Simulation Environment (MSE).

The remainder of this paper is organized as follows. In the next section we discuss related works. In Section 3 we describe the material and methods, which comprises the domain, student, tutor and interface modules. In Section 4, we discuss about the functioning of the ITS. In Section 5 is presented the conclusions and future work.

II. RELATED WORK

Several researches have been carried out to develop an architecture of an ITS which can capture some learner's information and use it to present an interactive and dynamic interface to the user.

Many different techniques to implement an ITS have been used, including the use of ontology to represent the knowledge, the apprentices characteristics [9] and the teaching strategies [10].

Some techniques to develop an Expert System were used to acquire the Learning Styles [11] of the apprentice to be stored in the Student Module and used as a parameter for the presentation of the content in the interface [12][13].

Tutor Module controls all the system and an Artificial Neural Network type Interactive Activation and Competition (ANN IAC) was used to implement this Module [12], which updates both the Interface Module and the Student Module [14]. An ITS can be integrated in a MSE which enables the learner to obtain experiences with a wider variety of structures and peculiarities, even being possible to repeat or redo procedures training free of charge [15].

The use of these architecture using tools, ontology, expert system and ANN IAC, including the Adaptive Hypermedia Systems shows up as a new approach to the development of ITS. Due this type of system involves various areas of knowledge applied with a common goal, the adaptation of the system in teaching to a particular student profile and its dynamicity is actually achieved with this approach.

III. MATERIAL AND METHODS

The first version of this ITS was developed as traditional architecture (Figure 1), which consists of four modules: Domain, Student, Tutor and Interface [14][16][17].

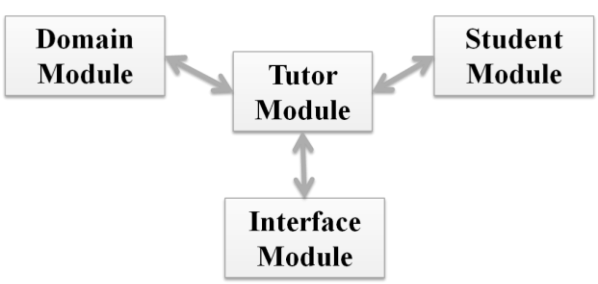


Figure 1. Architecture of the ITS.

A. Domain Module

The Domain Module was developed as an ontology model (Figure 2) that allows an easy way to organize and formalize the knowledge and ensure access with a single vocabulary. It can also represent organizational structures of a large complex domain and reason about itself [18].

The Protégé Software [19] from the Stanford University was used to build this Module. This software is based on Java and consists of a free open source ontology editor and knowledge-base framework that can be exported into a variety of formats including Resource Description Framework (RDF), Web Ontology Language (OWL) and Extensive Markup Language (XML) Schema.

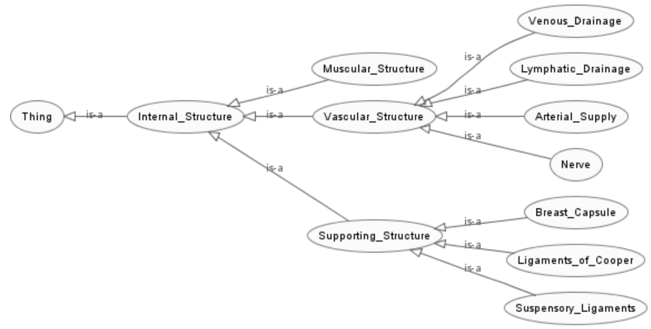


Figure 2. Breast Ontology (Protégé)

The idea of use ontology came from the fact that is needed to represent multimedia content as well as the medical vocabulary associated with the anatomy of the female breast and it's most important pathologies, including several types of breast cancer and other anomalies.

The Domain Module will include all the elements needed to represent the knowledge as well as a mechanism for automatic information retrieval (Figure 3). The ontology model is named ONTOMAMA-Model and the computational mechanism for information retrieval is known as ONTOMAMA-Engine.

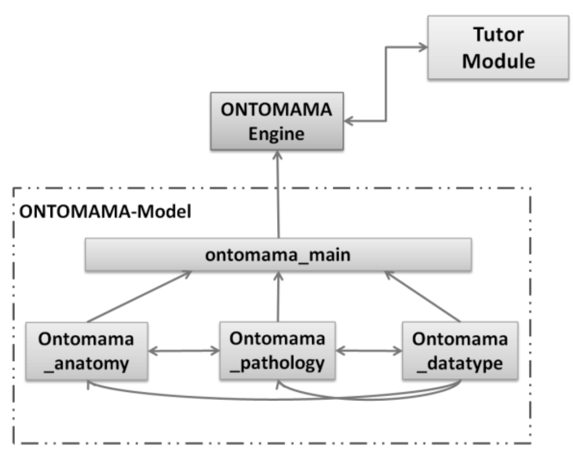


Figure 3. Ontomama Model.

The ONTOMAMA-Model is comprised by four distinct and interconnected OWL files (ontomama_main.owl, ontomama_anatomy.owl, ontomama_pathology.owl and ontomama_datatype.owl). The ontology was segmented to favor reuse and understandability of the model. While the file ontomama_main.owl will contain the model instances and their associated multimedia contents, the other three files will represent the medical vocabulary that depicts the anatomy and pathology of the female breast.

The ONTOMAMA-Engine has been programmed in JAVA by using Protégé API and Web Service technology. From this, it is intended to provide services in open standard format that might be consumed to retrieve any information in the ONTOMAMA-Model.

The challenge of the Domain Module of this ITS was to provide an efficient mechanism of tutelage where the requested of the apprentice can be intelligently delivered in the interface.

B. *Student Module*

The first version of the Student Module was developed as an Expert System (ES) – a system in which is possible to simulate the behavior of an specialist before a given situation - using the shell Expert Sinta [12][20]. This shell, developed in Delphi, uses a knowledge representation module which is based on production rules and probabilities and has an easy way to build screens and menus and a good inference engine shared. Because there is difficulty in integrating systems developed in Delphi in conjunction with other developed in Java, it was necessary to change the used shell.

The second version of this ES (Figure 4) was developed on Java Expert System Shell (JESS), a rule engine and scripting environment written entirely in Sun's Java language by Ernest Friedman-Hill at Sandia National Laboratories [21]. Similarly to Expert Sinta, JESS has the capacity to reason using knowledge supplied in the form of declarative rules. It is one of the fastest rule engines available, using an algorithm called Rete, which consists of a very efficient mechanism that organizes the rules as a tree, where similar rules are grouped at the same branches [22].

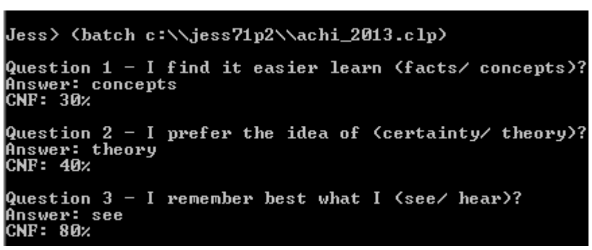


Figure 4. JESS Interface - Example of questions (CNF: degree of certainty of the learner's response).

The JESS Interface is not very friendly, which may cause the learner does not want to answer the questionnaire. To solve this problem a friendly and more attractive interface was created (Figure 5), so that the user does not feel unmotivated to answer the questions.

The Student Module is subdivided into two sub-modules. The first one is used to apply the initial questionnaire to the apprentice to identify his learning styles; this ES is intended to obtain qualitative information about the learner to enable the system to make decisions about the interface. The second one is activated when the user choice some different way to the interface present the content. Thereby, this modification caused by the user will both update the Student Module and the Interface Module.

Figure 5. JESS Interface.

The Student Module stores information about the apprentice's characteristics in a database (Figure 6) to facilitate the access of the needed information.

Name	Age	Sex	Carrer	Time Experience (years)	Learning Styles (%)	
Victor	53	Male	Teacher	25	Active	35
					Reflexive	15
					Intuitive	40
					Sensitive	75
					Verbal	80
					Visual	70
					Sequential	30
					Global	60

Figure 6. Student Module Database.

C. *Tutor Module*

The Tutor Module was developed by a mechanism of parallel distributed processing known as ANN IAC [23]. This topology is composed by units of processing (neurons) organized in groups that represent similar concepts, presenting characteristics of Bidirectional Associative Memories (BAM) [24].

This Module is responsible for all the behavior of the system, choosing what and how to present the content on the interface. For this purpose, the ITS will make use of the Student and Domain Modules that will provide information about the apprentice's behavior and the content used during the learning activities. From this, the ITS will adapt itself according to the individual profile and characteristics of each student.

D. *Interface Module*

The Interface Module was developed as an Adaptive Hypermedia System (AHS), which builds a model for each user based on their desires, preferences and knowledge stored on the Student Module and applies it to adapt to various visible aspects to the system. These systems work the concepts for the construction of hyperdocuments for presentation and navigation to suit to the user's needs [25].

IV. FUNCTIONING OF THE ITS

The first step is to login and create a password to the ITS. When it is done, the user must have to answer some questions about him (name, sex, age, career, time experience and a questionnaire to acquire his Learning Styles). This questionnaire consists of 22 questions, represented by 44 declarative rules described in the ES developed at JESS. Once the user has answered the questionnaire, all the information will be stored in a database which will serve as a parameter to the Tutor Module decides what is need to do in some occasions.

The second step is to choice the content that will be presented at the interface. Once the user has chosen the content, the Tutor Module will send to the interface the correct content in the correct type of media, according to the learning styles of the student stored on the Student Module. The Domain Module delivers multimedia contents to the Tutor Module which includes text, pictures, medical photographs, videos and animations. All of these to enrich the apprentice learning experience throughout a VR environment provided by the Interface Module.

The Tutor Module have as input values the apprentice's characteristics, that will be supplied by the Student Module; as output, the ANN IAC will have values related to the attributes (e.g., medias, videos, text, icons, images) stored in the Domain Module [14]. By means of the interactive activation feature, the ANN IAC can have as input values related to the attributes – actions of the apprentice – that will be back-propagated to retrieve values matching the apprentice's characteristics.

Due to this work in a bidirectional way, the network can both update the Interface Module by processing the apprentice's characteristics and the Student Module by processing the apprentice's actions in the Interface environment, thereby, representing the Didactic Ergonomic Knowledge (Figure 7) [7].

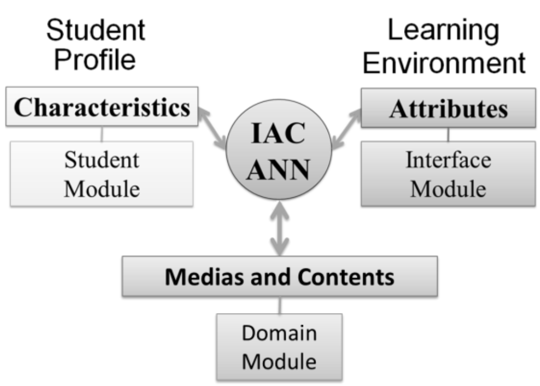


Figure 7. ANN IAC operation: update the learning environment according to the student characteristics and update the student profile according to the student actions in the learning environment, selecting the correct media type to be shown on the interface.

Besides retrieval by apprentice's characteristics and actions, ANN IAC also allows assignment of plausible default values in case of missing information and generalization over a set of instance units.

The Interface Module will present the attributes which will be related to the characteristics detected by the Student Module. The environment will supply icons, bottoms, media, text and images, in short, the resources available to the user to get a better preview of the content. If the learner wishes to define a different interface, modifying the type of media presented, he/she will have autonomy to select other way to represent the content. In this case, the Student Module will be updated storing the new characteristics related to this new media presented on the interface; and the Interface Module will be modified for the next screen there will be presented the content with media related to this updated done on the Student Module.

As mentioned before, once the questionnaire of the Student Module is correctly answered and converted to numeric values, representing the student profile, it will be an input to the ANN IAC. This input will create/ activate the student profile and this respective value will be inserted in the Ontology by the ONTOMAMA-Engine. The ONTOMAMA-Engine will choice the respective media of the content supplied by the Domain Module, and it will be given back to the ANN IAC. The Tutor Module will, then, show on the Interface the content adapted to the student profile.

Therefore, the validation process of this ITS is in progress because the class period has not yet started. Three Universities will be participating in this process: the University of Brasília (federal) and the Catholic University of Brasília (particular), both located in Brasília city in Brazil, and the Universidad de La Frontera (federal) located in Temuco in Chile.

V. CONCLUSION AND FUTURE WORK

This work has a great relevance on the scope of Education and Health areas, since this system helps in the teaching learning process of the female breast.

The architecture proposed in this paper proves to be ideal in order to submit a facility in the organization and changing the knowledge, including the sharing of information with others systems, through the use of ontologies.

The identification of the profile of the learner through the ES proved to be very effective and this information, used to determine the media which is going to be displayed in the interface, was quite consistent with the user request.

The use of an ANN IAC in this ITS was very efficient due its capacity to associate the apprentice's characteristics to the contents and its dynamics process of updating the media on the interface and the profile of the learner, during the use of the system.

Intended future works refer to: a) integrate this ITS in a MSE based on VR, in which will be constructed 3D models using real images and an interaction based on the touch of organs and others structures to make possible the embodiment of invasive medical procedures related to the study of the female breast. This tactile interaction will use an Haptic Interface [26][27] called Omega 7 [28], a dispositive that works together with a graphic interface designed to promote a tactile simulation; b) to improve the development of the Student Module terms of access to information in the database and providing this information to the Tutor Module by using fuzzy logic techniques.

ACKNOWLEDGMENT

The authors thanks to CAPES and PQ 2F/ CNPq for finance and support research grant and acknowledge INCT/ MACC and Ministry of Health by the financial support to the development of this research.

REFERENCES

[1] Carbonell, J. R. AI in CAI: an artificial intelligence approach to computer assited instruction. IEEE Transactions on Man Machine Systems, v.11, n.4, pp. 190-202, 1970.

[2] GLOBOCAN. Section of Cancer Information (IARC), 2008. Available at: http://globocan.iarc.fr. Acessed on January 29, 2013.

[3] INCA. National Cancer Institute of José de Alencar Gomes da Silva/ Healthy Ministry, 2011. Estimative 2012: Incidence of Cancer in Brazil. Rio de Janeiro (In Portuguese).

[4] Curilem, G. M. S, de Azevedo, F. M and Ferneda, E. Artificial Intelligence in Education Support Health. In BRASIL, L. M. (Ed.). In Health Informatics. Taguatinga e Londrina: Universa e Eduel, v. 1, pp. 255-271, 2008 (In Portuguese).

[5] Urretavizcaya, M. L. Sistemas Inteligentes em el âmbito de la educación. Iberoamericana Journal of Artificial Intelligence, v. 5, n. 12, pp. 5-12, 2001.

[6] Thiry, M. C. C. An Agent Based Architecture to Support Distance Learning. Florianópolis. Doctoral Dissertation, Federal University of Santa Catarina, 2001 (In Portuguese).

[7] Curilem, G. M. S. and De Azevedo, F. M. Didact Ergonomy for the Interface of Intelligent Tutoring Systems. In Computers and Education: Towards a Lifelong Learning Society Kluwer Academic Publishers. Netherlands, pp.75-88, 2003.

[8] Melo, J. S. S., Brasil, L. M., Panerai, C. E. B. and da Silva, A. P. B. Integration Interface Phantom of the Intelligent Tutoring System for Medical Simulation Environment. Brasilian Journal of Biomedical Engineering, v. 27, n. 2, pp. 98-109, 2011 (In Portuguese).

[9] Latham, A. M., Crockett, K. A., McLean, D. A., Edmonds, B. and O'Shea, K. Oscar: an Intelligent Conversational Agent Tutor to Estimate Learning Styles. IEEE International Conference on Fuzzy Systems (FUZZ), 2010.

[10] Zhang, Z., Geng, X., Jiang,Y. and Yang, Y. An Intelligent Tutoring System (ITS) for Tactical Training based on Ontology. IEEE International Conference on Information Engineering and Computer Science (ICIECS), 2009.

[11] Felder, R.M. and Solomon, B. A. Index of Learning Styles Questionnaire. Available at: http://www.engr.ncsu.edu/learningstyles/ilsweb.html, Accessed on January 29, 2013.

[12] Curilem, G. M. S. Methodology for The Construction of Adaptive Interfaces for Intelligent Tutoring Systems. Doctoral Dissertation. Federal University of Santa Catarina, 2002.

[13] Zatarain-Cabada, R., Barrón-Estrada, M. L., Zepeda-Sánchez, L., Sandoval, G., Osorio-Velasquez, J. M. and Urias-Barrientos, J. E. Modeling student's learning styles in web 2.0 learning systems. World Journal on Education Technology – Special Issue: Artificial Intelligence on Education -, v.1, n.2, pp. 78-88, 2009.

[14] Curilem, G. M. S., Barbosa, A. R. and de Azevedo, F. M. Intelligent tutoring systems: Formalization as automata and interface design using neural networks. ScienceDirect. Computers & Education v. 49, issue 3, pp. 545-561, 2007.

[15] Melo, J. S. S., Brasil, L. M., Ferneda, E., Balaniuk, R., Costa, E. de B., Bitencourt, I. and Rocha, L.Use of Virtual Reality in Intelligent Tutoring Systems Intended for Teaching Anatomy. XVIII Brazilian Simposium on Computing Education (SBIE). Mackenzie, pp. 51-54, 2007 (In Portuguese).

[16] Wenger, E. Artificial Intelligence and Tutoring Systems. Los Altos, CA: Morgan Kaufmann, 1987.

[17] Freedman, R. What is an Intelligent Tutoring System? Published in Magazine Intelligence, v.11, issue 3, pp. 15-16, 2000.

[18] Lopes, L. F. A model of engineerting knowledge-based ontology and probabilistic calculation support for diagnosis. Doctoral Dissertation. Federal University of Santa Catarina, 2011 (In Portuguese).

[19] Protege. The National Center for Biomedical Ontology. Stanford Center for Biomedical Informatics Research. Available at: http://protege.stanford.edu/. Accessed on January 29, 2013.

[20] LIA. Artificial Intelligence Laboratory of Federal University of Ceara. Fortaleza, Brazil. Available at: http://www.lia.ufc.br/site/. Accessed on January 29, 2013.

[21] JESS. Rule Engine For The Java. Sandia National Laboratories. Livermore, CA. Available at: http://www.jessrules.com/. Accessed on January 29, 2013.

[22] Friedmann-Hill, E. JESS in Action: Ruled-Based System in Java. USA: Manning Publications, 2003.

[23] McClelland, J. L. and Rumelhart, D. E. Explorations in Distributed Processing: A handbook of models, programs and exercises. Ed. Bradford Book. Massachusetts Institute of Technology, 1989.

[24] De Azevedo, F. M., Brasil, L. M. and Oliveira, R. C. I. Neural networks with applications in control and expert systems. Florianópolis: Visual Books, 2000 (In Portuguese).

[25] Fragelli, R. R. An Approach of Quantized Network and Multiform Objects to Model Domains in Intelligent Tutoring System. Doctoral Dissertation. University of Brasilia, 2010 (In Portuguese).

[26] Strauss, E. Feeling the Future. In: Scientific American Presents – Your New Senses, Scientific American Digital, 1999.

[27] Mauerberg-deCastro, E., De Paula, A. I., Tavares, C. P. and Moraes, R. Space Orientation for Visually Impaired Adults: Effects of a Training Navigation. Psychology: Reflection and Critique, v. 17, n.2, pp. 199-210, 2004 (In Portuguese).

[28] Force Dimension. Force Dimension Haptic Devices. Available at: http://www.forcedimension.com/. Accessed on January 29, 2013.

CyPhy-UI: Cyber-Physical User Interaction Paradigm to Control Networked Appliances with Augmented Reality

Through Design, Implementation, and Evaluation of EVANS
(Embodied Visualization with Augmented-Reality for Networked Systems)

Kenya Sato, Naoya Sakamoto, Shinya Mihara, and Hideki Shimada
Department of Information Systems Design
Doshisha University
Kyoto, 610-0321 Japan
e-mail: ksato@mail.doshisha.ac.jp, pooh708s@gmail.com, dezeckco@gmail.com, hshimada@mail.doshisha.ac.jp

***Abstract*—Many kinds of networked home appliances, which are connected by standardized control functions, have recently appeared and continue to increase. Because a general infrared remote control is for single-way communication from a remote control to a specific appliance, but not to receive signals from an appliance to a remote control, it is impossible to gather an appliance's information with a remote control. Meanwhile, unlike a general infrared remote control, it is difficult to control a specific appliance because users can simultaneously operate all of their appliances with a WiFi controller. In this paper, to control networked appliances with a smart phone or tablet computer as a WiFi controller, we propose a new interface paradigm called a cyber-physical user interaction that creates virtual (cyber) space, sends commands, and receives responses from networked (physical) appliances through space with augmented-reality (AR) technology. With the paradigm, which enables interconnectivity among appliances from various vendors, it is possible users with uniform and intuitive operation of home appliances. In addition, we implement and evaluate an Embodied Visualization with Augmented-Reality for Networked Systems (EVANS) that controls a system of home appliances and sensor devices through a cyber-physical user interaction (CyPhy-UI) paradigm by a web camera to retrieve information from real world environments and touch-screen display to show AR visualization and user interaction components to retrieve user input.**

Keywords-appliance; control; cyber-physical; user interface; augmented reality

I. INTRODUCTION

The appearance and popularity in homes of such audio/video appliances, as TVs, DVD players, hard-disk recorders, audio receivers, and digital speakers, continue to increase. Recently these appliances can be connected to each other through home networks with standardized network functions (e.g., DLNA [1]) to transmit audio/video stream data. In the near future, we expect that household electrical appliances, air conditioners, floor lamps, electric curtains, and sensor devices (for temperature, humidity, or illumination) will also be connected to home networks to enhance convenience in *smart homes* [2][3]. In general, since the controls over home appliances are used from exclusive infrared remote controls, as the number of home appliances increases, distinction by remote controls becomes more complicated. Because a general infrared remote control is for single-way communication (to send signals from a remote control to an appliance, but not to receive them from an appliance to a remote control), it is impossible gather an appliance's information with a remote control. Meanwhile, unlike a general infrared remote control, it is difficult to control a specific appliance because users can simultaneously operate all of their appliances with a WiFi controller.

In this paper, to control networked appliances with a smart phone or a tablet computer as a WiFi controller, we propose a new user interaction scheme called a cyber-physical user interaction (CyPhy-UI) paradigm in which virtual (cyber) space is created to send commands and receive responses from networked (physical) appliances through space with augmented-reality (AR) technology.

II. CYBER-PHYSICAL USER INTERACTION PARADIGM

A. Categorization of Control Types for Appliances

To analyze the control methods for networked appliances, we need to categorize user interaction models among a user and appliances. Suppose several appliances around a user, who chooses one of them and controls it by remote control. As described Figure 1, we categorized the situation into the following six types:

1) *Type 1—Direct command and direct response:* A user directly sends commands to Appliance A, and directly receives responses from it. For example, TVs are included in this type. When a user changes the channel with a remote control, she pushes a switch on it. The channel information is shown on the TV display, and she can know that follow the channel has been changed.

2) *Type 2—Direct command and no response:* A user directly sends commands to Appliance A without a response from it. For example, an air conditioner (without any display function on the appliance) is included in this type. When a user tries to turn down the temperature with a remote control, he pushes a switch on it. However, there is no direct response from the air conditioner. Although he may eventually feel that the temperature has been lowered, he

 ISBN: 978-1-61208-250-9

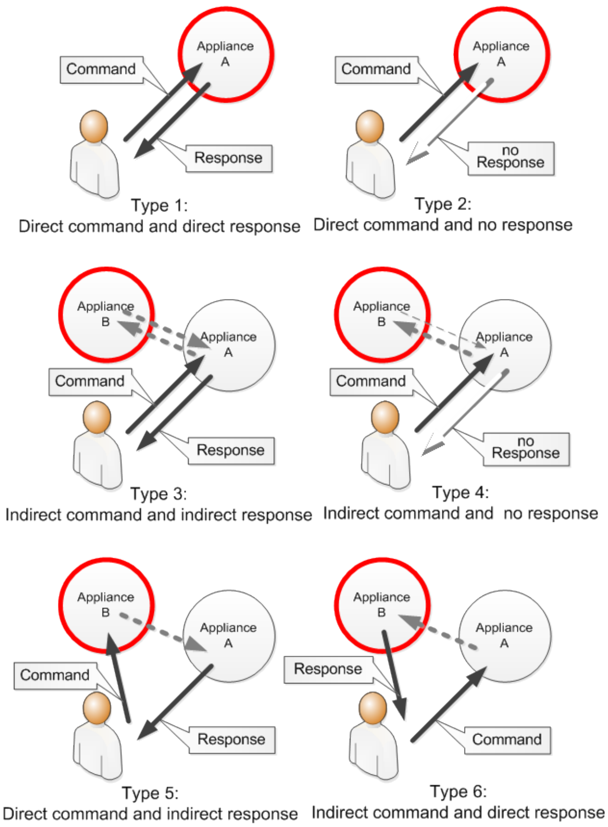


Figure 1. Six types of interactions among users and appliances.

cannot immediately check the air conditioner's status, because there is no response from Appliance A.

3) Type 3—Indirect command and no response: A user controls Appliance B through Appliance A without a response from Appliances A or B. For example, a speaker (without any display function) and an audio receiver are included in this type. When a user increases the volume of a speaker connected to an audio receiver with a remote control, he pushes a switch on it. The speaker becomes louder, but not the audio receiver. In addition, there is no response from the speaker or the audio receiver. A user may not directly understand which appliance to control.

4) Type 4—Indirect command and indirect response: A user controls Appliance B through Appliance A, and the response of Appliance B from Appliance A. For example, a TV set with separate speakers is included in this type. When a user increases a speaker connected to the TV set with its remote control, he pushes a switch on it. The speaker becomes louder, and the TV shows the volume status on its display. The user can realize that understand the volume has increased. However, suppose that six speakers (5.1 channels) are connected to the TV, and the user wants to decrease their volume. Identifying a certain one may be difficult without speaker identification on the TV display.

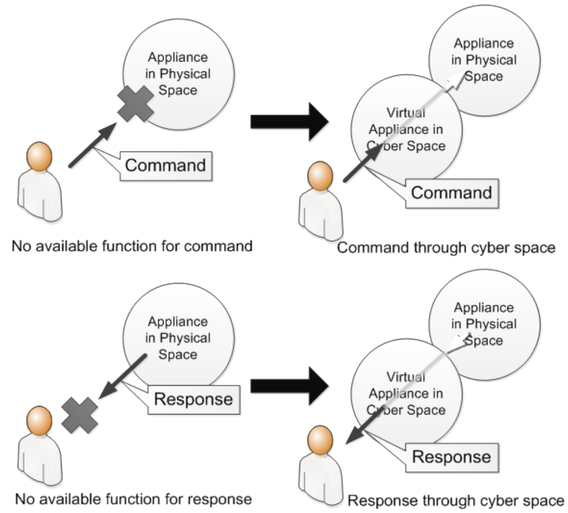


Figure 2. Request and response model with/without virtual appliances in cyber space.

5) Type 5—Direct command and indirect response: A user directly controls Appliance B, and its response from Appliance A. For example, a DVD recorder is connected to a TV, and its status is displayed on the TV's display. When a user wants to identify the track of a DVD recorder connected to the TV with DVD recorder's remote control, she pushes a switch on the remote control. The DVD's tracks are displayed, and the TV shows their status. She can identify specific tracks. Even if both a DVD recorder and a hard disk video recorder are connected to her TV, she can differentiate between them and choose which to control.

6) Type 6—Indirect command and direct response: A user controls Appliance B through Appliance A and direct responses from Appliance B. For example, a speaker with volume indicators and an audio receiver are included in this type. When a user increases the volume of a speaker connected to an audio receiver with a remote control, she pushes a switch on it. The speaker becomes louder, and she can check the indicators on the speaker. However, it is difficult to identify a certain appliance among several only with the information on the audio receiver.

Among these six types, Type 1 direct control and direct response is the ideal case to control appliances. With it, users can directly control a target appliance and realize their request. However, in real situations, an appliance may have no display function and/or may have no direct control method with a remote control.

In addition, users might get not only an appliance's status information but also audio/video contents from the appliance as response messages without sending a request to it.

B. Approach

To send a request from a user to an appliance without functions to receive commands, and/or to receive responses

 ISBN: 978-1-61208-250-9

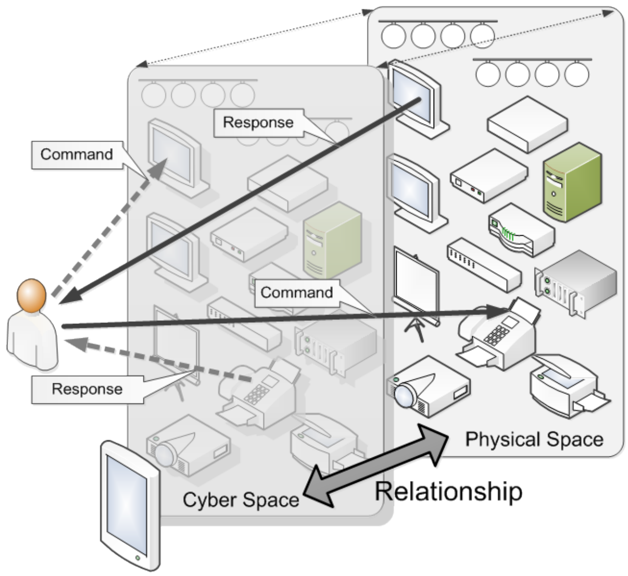


Figure 3. Relationship between appliances in cyber and physical space.

from an appliance to a user without functions to send a response, our approach creates a new paradigm. In it, a virtual appliance is defined in cyber space, and users can send a command to or receive responses from a virtual appliance in cyber space. Our cyber-physical user interaction (CyPhy-UI) paradigm is shown in Figure 2.

We apply such portable devices, as smart-phones and tablet computers with augmented reality technology to realize our CyPhy-UI paradigm. Portable devices have a touch display and a camera installed on its back, as well as a WiFi network function. With the camera, the appliances in the physical space are shown as camera images on the display, which means appliances in cyber space. The appliances in the physical space must be connected to the portable device. As described in Figure 3, a menu to control an appliance is shown on the portable device's display, and users can send a command by touching the menu. In addition, they can identify the appliance's responses, which are shown on the display.

III. Related Works

A. Tangible User Interfaces

Graphical user interfaces makes a fundamental distinction between input devices, such as the keyboard and mouse as control and output devices like monitors for the synthesis of visual representations. Tangible User Interfaces [4][5] proposed by Hiroshi Ishii et al., couple physical representations (e.g., spatially manipulable physical objects) with digital representation (e.g., graphics and audio), yielding interactive systems that are computationally mediated, but generally not identifiable as computers in itself. The design and selection of appropriate physical representations is an important aspect of tangible user interface design.

In our CyPhy-UI paradigm, virtual representations are defined in cyber space and users can send a command to or receive responses from physical objects through cyber space without using physical representations. In addition, virtual representations in CyPhy-UI paradigm are real camera images of physical objects, and users can easily identify physical objects and their functions.

B. Augmented Reality Systems

Many research projects make invisible information related to objects visible using augmented reality technology, such as Google's Project Glass [6], which is an outdoor mobile augmented reality street view application, MARA [7], which is a sensor-based augmented reality system for mobile imaging device. uMegane [8] which is a visualization system of sensor data with AR technology, easily acquire sensor data for users who are unfamiliar with sensor technology. Extate [9] is a visualization system of a wireless network with AR Technology that enables users to acquire such network status as packet data and network type. Sekai Camera[10] is a popular smart-phone application to enable users to view AR information about subjects of scenery. Our goal is not only to acquire appliance information but also to control sensor nodes and networks. Thus our target system is different from existing researches.

C. LED Visual Markers for Augmented Reality

In present AR technology, an image marker [11] is necessary to identify the AR graphics to display on the camera screen. However, some issues have presented, such as marker size, lighting environments, and distance from a controller to appliances. In related works using a LED for the AR marker, Visual Computer Communication (VCC) marker [12] was proposed. VCC markers use 16 LEDs as one AR marker, so cameras can receive the increased lighting information of LEDs, and display the AR graphics. However when operating in appliances, a VCC requires far more insertion space into appliances than our proposed LED visual marker. Since we only need the networked home appliance's ID information and the LED location displayed through the camera, few data are necessary. For this reason, VCC markers are not suitable for operating appliances. As described below, we propose an LED visual marker and a network home appliance operating system that intuitively and easily operates complex network appliances in dark, bright, or distant home environments.

IV. Realization of The New Paradigm

A. Design of EVANS

We designed and implemented EVANS, or Embodied Visualization with Augmented-Reality (AR) for Networked Systems, which is a system to provide users with a uniform and intuitive interface of home appliances and/or sensor networks using a CyPhy-UI paradigm.

AR technology can provide images generated by overlapping virtual information on real environmental

 ISBN: 978-1-61208-250-9

Figure 4. Using EVANS.

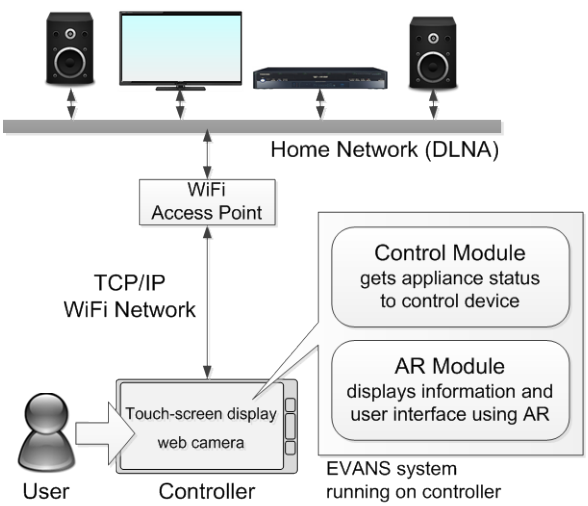


Figure 5. Implemented system architecture for EVANS.

Figure 6. Display image on a tablet computer for EVANS.

Figure 7. Volume control of right speaker with AR.

images captured by cameras. The virtual information in cyber space called an annotation includes the information associated with certain objects in real space. In general, AR technology uses an AR marker to detect the camera's position and orientation, which is two-dimensional code. In advance, the pattern data of AR markers are registered for the application, which can recognize the object in real space by tracking the AR markers with the camera.

For example, as shown in Figure 4, the user points the camera, which is installed on the back of the remote control (e.g., smart-phone or tablet computer) at the desired appliance, whose control interface for the specific appliance is shown on the controller's front display. To create annotations on the display, a specific AR marker is attached to each appliance. When the right speaker is chosen, the volume up/down control menu is shown as annotation information (Figure 5). The user then simply changes the volume. If a TV is chosen, appliances connected to it are shown; if a DVD recorder is chosen, the playback controls or content selection are shown on the controller's display. The user simply performs the operations. In addition, by providing a uniform user interaction, users can perform operations on any appliance connected to their home network.

Building upon standardized home networks that enable interconnectivity among appliances from various vendors, we aim to solve the appliance selection problem by providing a real world display augmented with appliance information and control interfaces.

B. Implementation of EVANS for AV Appliances

The system architecture for implementing EVANS to control and monitor AV appliances is shown in Figure 5. We used a web camera to retrieve information of the appliances in the real space, and a touch screen display to show AR visualizations and user interaction components and to retrieve user input. The entire system consists of AR and control modules.

The AR module displays appliance information and user interaction components after the web camera is pointed at an appliance. It also gathers input from the user and presents the intended control data to the control module. For example, the user may point the web camera at a speaker with a certain identifying marker. The control module communicates with

TABLE I. DMS RESPONSE TIME FOR ORIGINAL DLAN FUNCTION AND EVANS

Evaluation item	***Original DLNA function***	***EVANS***
Response time (msec)	0.33	0.39

the speaker to retrieve its volume value. The AR module places the user interaction components (in this case a volume value display and volume control buttons) on the marker. When the user operates the buttons, the AR module gives the control data to the control module, which sends the commands to the appliance, and the AR module updates the AR display.

This module communicates with such home appliances as TVs, HDD recorders and speakers and actually controls them based on user input and retrieves information about them. The control module uses the Digital Living Network Alliance (DLNA) to communicate with them.

We implemented EVANS using a note PC and a tablet computer. The note PC for the home appliances and the tablet computer were connected to the network, and to run DLNA Media Server applications (TVersity [13] and DiXiM [14]) that act as a DLNA Media Server (DMS) to provide video content. External speakers were attached to the note PC. The node PC and the speakers have identifying markers for use in AR. The following is the implementation setup:

- AR display: ARToolKit [15]
- DLNA controls: Cyberlink for C++ [16]
- Multimedia display: Simple DirectMedia Layer (SDL) [17]
- Graphics library: Freetype Graphic Library (FTGL)

Figure 6 shows the state of the AR display when the search button is pressed. It confirms that the appliance information was successfully discovered on the network; the names of the DMS applications and the details of the stored content for the note PC are overlaid on the marker by AR. Intuitive user interaction components are also displayed for the speakers attached to the note PC; the volume values and control buttons are displayed by AR. Figure 7 shows the AR displays when the volume is set to 50. We confirmed that user input against the system actually changed the volume of the speakers and that the AR display was updated accordingly.

To evaluate our implementation's practicality, we measured the time before a response from the DMS is received after issuing a search for it. As a control value, we also measured the same response time for existing DLNA appliances. Table 1 shows the average results for 20 individual measurements. Since our implemented EVANS causes no major delay compared to the existing system, it is practical.

C. Implementation of EVANS for Sensor Networks

Sensor networks are currently used in such fields as home electronics, energy management, and security because

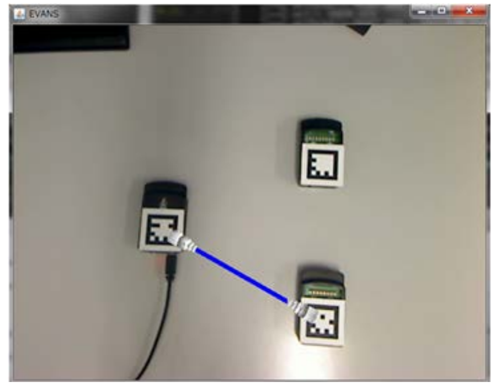


Figure 8. Displayed network connection information among sensor nodes.

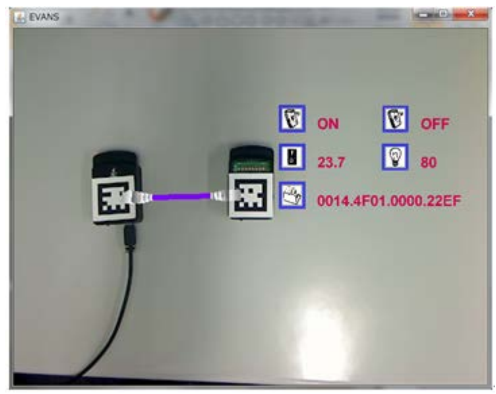


Figure 9. Displayed sensor information included on nodes.

they can immediately detect an event or a situation and automatically control an actuator. However, managing each sensor node connected to wireless sensor networks is difficult because sensor node status and wireless network topology are invisible. EVANS can display sensor data and such network information as the link status, the packet data, and the traffic in the sensor network as AR information in cyber space on the display.

We implemented EVANS [18] for a sensor network using Sun Small Programmable Object (SunSPOT) [19] as the sensor node and a Java version ARToolKit (NyARToolkit [20]) to generate AR images. SunSPOT, which is a wireless sensor network device, can measure temperature and illuminance and is also equipped with a push button switch.

Figures 8 and 9 show link status images of our prototype system captured by the sensor node's camera. In real space images, we cannot directly see the connection and the sensor information between sensor nodes, but this system allows us to directly acquire such information through AR images.

This system has two operation methods. One shows the resource information. When users tap a sensor node on the touch screen, resource information is displayed. Another controls the connection between sensor nodes. When users drag and drop between sensor nodes, the switches of both nodes are toggled. In the installed program, if the switch of each node is turned on, these nodes communicate with other sensor nodes. If the switch of each node is turned off, the connection of these nodes is interrupted. When the connection is interrupted, the virtual link cable of the AR annotation image disappears.

V. DISCUSSIONS

A. Current Issues with AR Solution

Presently, infrared remote controllers are the most popular way of controlling home appliances. Although some full-feature remote controllers allow users to control multiple appliances, in most cases each appliance comes with its own remote controller. These remote controllers send out infrared signals to control their appliances, but it is generally not possible to conversely retrieve appliance status information at the controller. This is partly compensated by displaying status information on the appliance itself if the target appliance is a television or is connected to a television, but most other appliances that are not equipped with a display only provide limited status information (most commonly using LEDs). Also, such additional detailed information as instruction manuals cannot be used directly on remote controllers.

Most people have at least one mobile phone and carry it when they are at homes [21], and smart-phones are also becoming more and more popular. Recent home appliances, especially multimedia appliances, can be easily connected to a local network. Based on these circumstances, traditional remote controllers are unnecessary if these smart phones could control home appliances. Furthermore, through smart-phone display, users can learn detailed information about their appliances and control them by touch-screen, providing further convenience.

ARToolKit, a popular AR framework also used in our above implemented system, uses image markers or graphic patterns that are usually printed on paper to identify objects and their placements. Figure 10 shows an example image marker used in ARToolKit. This framework poses a number of problems when applied to home appliance control. To address these problems, we proposed LED visible markers [21], where LEDs equipped on appliances blink at a fast rate and are used as AR identification markers (Figure 11). The following sections discuss the issues with image markers and the solutions that can be provided by LED visible markers.

B. Visual Attractiveness

Image markers tend to be rather large because the camera must recognize their patterns from a certain distance. They also need to be attached in places where they are always visible. These constraints make the appliances visually unattractive. LED visual markers, on the other hand, are hardly noticeable. They are also practical because the LEDs

Figure 10. Image marker for ARToolKit..

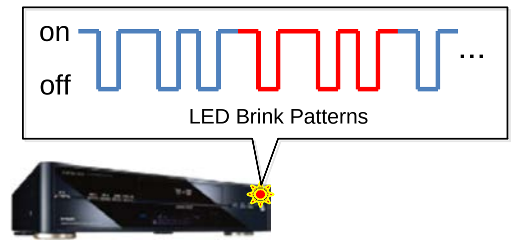


Figure 11. LED visible marker.

of most home appliances are lit up even when their power is turned off.

C. Recognition

Because home appliances are almost always used indoors, they must be controlled even when the room is dark. Image markers basically require that the room is lit up for them to be recognizable, but the LED visual markers are self-luminous and recognizable in complete darkness. Note that this does not imply that they are less recognizable in brighter rooms.

D. Dynamic Identification Changes

Because image markers are generally printed on paper, changing an appliance ID is cumbersome. Another maker must be printed to replace the old one. On the other hand, LED Visible Markers allow dynamic changes of appliance IDs. They are also more useful in networked setups because one could possibly use an appliance's MAC or IP address as part of the appliance ID.

E. Directional Recognition

One benefit of image markers is that they are two-dimensional, which lets the camera compute the direction and distance of the marker. LED visual markers, on the other hand, are zero-dimensional points, and do not yield such information. As a countermeasure, three LEDs of different colors can be placed in a triangular shape. For applications in appliance controls, since user interaction components merely need to be placed vertically and never in other directions depending on the marker position, this may not be an issue at all.

ISBN: 978-1-61208-250-9

F. Control of Legacy Appliances

Some legacy appliances may not have network control functions and can be controlled only with an infrared remote control. In this case, we use an image marker stuck to an appliance instead of a battery-powered LED visual marker to identify it. We also developed remote control software on a smart-phone through iRemocon [23] (universal infrared remote control device connected to PC) to send control signal to an appliance.

VI. Conclusion and Future Work

The popularity of home appliances that can interconnect with other home appliances through networks continues to increase. The operation and function of such appliances are complex since they can share contents and data with other network home appliances. However, it remains difficult for users to identify network home appliances, since their locations are not installed and cannot be displayed easily when operating them. Because of these problems, more obvious and intuitive operations are needed for controlling networked home appliances.

We proposed a new user interaction paradigm to control networked appliances. In our cyber-physical user interaction (CyPhy-UI) paradigm, cyber space is created and we can send commands to and receive information from physical appliances through it with augmented reality technology. With our paradigm, which enables interconnectivity among appliances from various vendors, we can provide users with uniform and intuitive operation of home appliances.

We also implemented an Embodied Visualization with Augmented Reality for Networked Systems (EVANS), which controls a system of home appliances and sensor devices through a CyPhy-UI paradigm, using a web camera to retrieve information from real world environments, and a touch screen display to show AR visualizations and user interaction components and to retrieve user input. We also evaluated our system using this method based on response times, and conclude that since our system does not introduce noticeable response delays, it is practical. We also discussed the issues about current AR image markers when applied to home appliance control, and discussed the practicality of LED visual markers to solve them.

Future research will implement the LED visual marker method and evaluate such characteristics as recognition rates. We also will produce a more general and practical implementation using handheld devices with limited computational resources such as smart phones.

Acknowledgment

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (KAKENHI) Grant Number 21500084.

References

[1] DLNA: Digital Living Network Alliance, DLNA Networked Device Interoperability Guidelines, expanded, 2006.

[2] Scott Davidoff, Min Kyung Lee, Charles Yiu, John Zimmerman, and Anind K. Dey, Principles of Smart Home Control, Ubicomp 2006 LNCS Vol.4206, pp.19-34, 2006.

[3] Richard Harper, FromSmart Home to Connected Home, The Connected Home: The Future of Domestic Life, Part 1, pp.3-18, 2011.

[4] Hiroshi Ishii and Brygg Ullmer, Tangible bits: towards seamless interfaces between people, bits and atoms, Proceedings of the SIGCHI conference on Human factors n computing systems, pp.234-241, 1997.

[5] Brygg Ullmer and Hiroshi Ishii, Emerging frameworks for tangible user interfaces, IBM Systems Journal, Vol.39, Issue 3/4, pp.915-931, 2000.

[6] Google, Project Glass, Thougts, designs, and stories, http://www.google.com/+projectglass/ [Jan. 1, 2013].

[7] Markus Kähäri and David J. Murphy, MARA-Sensor based Augmented Reality System for Mobile Imaging Device, 5th IEEE and ACM International Symposium on Mixed and Augmented Reality, 2006.

[8] Takuya Imaeda, Kazunori Takashio, and Hideyuki Tokuda, uMegane: Visualization System of Sensor Data using AR Technology, IPSJ Technical Report Vol.2008, No.66, pp.39-44, 2008.

[9] Takuya Takimoto, Namatame Naoya, Jin Nakazawa, Kazunori Takashio, and Hideyuki Tokuda, Extate: Visualizing Wireless Networks by using AR Technology, The 3rd Asia-Europe Workshop on Ubiquitous Computing, 2010.

[10] Sekai Camera Beyond Reality, http://sekaicamera.com/ [Jan. 1, 2013].

[11] Hirokazu Kato and Mark Billinghurst, Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System, Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, pp.85-94, 1999.

[12] Yasue Kishino, Masahiko Tsukamoto, Yutaka Sakane, and Shojiro Nishio, Realizing a Visual Marker Using LEDs for Wearable Computing Environment, Proceedings of IEEE International Conference on Distributed Computing Systems Workshops (IWSAWC2003), pp.314-319, 2003.

[13] TVersity, the Leading UPnP & DLNA Media Server, http://tversity.com/ [Jan. 1, 2013].

[14] DigiOn, DiXiM Digital TV, http://www.digion.com/pro/dxtv/ [Jan. 1, 2013].

[15] Hirokazu Kato, ARToolKit: Library for Vision-based Augmented Reality, IEIC Technical Repor, Vol.101, No.652, pp.222-232, 2002.

[16] QNX, Simple Directmedia Layer, http://www.libsdl.org/ [Jan. 1, 2013].

[17] Cybergarage, CyberLink for C++ , http://www.cybergarage.org/twiki/bin/view/Main/CyberLinkForCC/ [Jan. 1, 2013].

[18] Naoya Sakamoto, Hideki Shimada, and Kenya Sato, Design and Implementation of Sensor Network Device Control System with AR Technology, Advanced Engineering Forum, Vols.2/3, pp.131-134, 2011.

[19] Oracle Labs, Sun SPOT World Home, http://www.sunspotworld.com/ [Jan. 1, 2013].

[20] NyARToolkit project, http://nyatla.jp/nyartoolkit/wp/ [Jan. 1, 2013]..

[21] TechBed.com, Thigns To Do With A Mobile Phone, http://www.techbed.com/things-to-do-with-a-mobile-phone/ [Jan. 1, 2013]..

[22] Shinya Mihara, Akira Sakamoto, Hideki Shimada, and Kenya Sato, Augmented Reality Marker for Operating Home Appliances, Proceedings of the 2011 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing (EUC2011), pp.372-377, 2011.

[23] iRemocon, http://i-remocon.com/ [Jan. 1, 2013]

Luminance Contrast Influences Reaction Time in Young and Older Adults

Patrick J. Grabowski and Andrea H. Mason
Department of Kinesiology
University of Wisconsin-Madison
Madison, WI, U.S.A.
pgrabowski@wisc.edu, amason@education.wisc.edu

***Abstract*—Age-specific design principles for three dimensional virtual environment systems are sparse. Given that sensorimotor control systems change across the lifespan, understanding age differences in motor performance within virtual environments is crucial to designing effective, usable interfaces. This paper investigates the effect of luminance contrast level on reaction time to a visual stimulus in both young and senior adults. Results indicate that young adults have faster reaction times than seniors, but both groups improved reaction times with increasing luminance contrast of the target. Young adults improved at lower levels of contrast than seniors. Implications for age-specific design of virtual environments are discussed.**

Keywords- virtual environment; aging; motor control; reach to grasp; luminance contrast

I. INTRODUCTION

The aging of our population presents a number of challenges for the coming decades. In particular, aging brings about potential loss of an individual's function due to disease, injury, or the degenerative nature of aging itself. This results in significant burden on caregivers, healthcare systems, and economies. For example, the aging of the United States' population is the driving factor behind a predicted 300% increase in medical spending on cardiovascular diseases (i.e., coronary heart disease and stroke, among others) by the year 2030 [1]. One potential intervention for performance enhancement and rehabilitation of elderly individuals involves the development of computer-based technologies designed as adjuncts to traditional intervention methods. Specifically, three-dimensional (3D) virtual environments (VE) have been identified as systems with good potential to serve in these types of applications [2, 3].

The ultimate purpose of VEs is to provide the user with a computer-based tool in which a variety of common and novel manipulative, cognitive, and exploratory activities can be performed easily and efficiently. Applications of such technology often target younger users, such as surgical training and flight simulation [4, 5]. A much smaller number of applications address special populations of elderly. For example, VEs designed in various formats assist individuals in rehabilitation after stroke, including training of obstacle avoidance during walking, and enhancement of driving skills [6-8].

While the application of this technology broadens, authors continue to point out distinct weaknesses of virtual environments [2, 3, 9]. First, cost-effectiveness remains problematic, as many systems employ extensive computing resources and require technical expertise to run. VEs also typically utilize complex visual graphics, a trait inherent in the design of programs targeting young users. This increases the overall cost of production, and these visually rich displays also have the potential to be overly distracting, hampering motor performance in older individuals and producing discomfort or even nausea (i.e., cybersickness) [3]. While cost and comfort remain problematic, the foremost underlying barrier to implementation of VEs into real-world applications for the general population continues to be an incomplete understanding of the human sensorimotor system [10, 11]. When one considers that this system changes across the lifespan, this barrier becomes further complicated. Indeed, the sensorimotor performance of older individuals is not equivalent to that of the younger users targeted in most 3D computer applications**.** Therefore, the purpose of this paper is to provide information about the age-associated use of specific visual information in virtual environments. Background information will be presented, followed by a description of the experimental methodology. Results will then be reviewed followed by a discussion of their significance and relevance to future work.

A. Human computer interaction across the lifespan in 3D environments

Evidence of age-related differences in performance between young and older adults indicates disparities in reactions to environmental immersion, usage of input devices, size estimation ability, and navigational skills (for review see [12]). While these studies indicate a need for age-specific design principles, very little knowledge regarding sensorimotor control in VEs exists. Currently, the International Encyclopedia of Ergonomics and Human Factors leaves the explanation of age-related differences in virtual environments to a two sentence description recommending that equipment be tailored to physically fit the smaller frames of children, and for designers to take into consideration the changes in sensory and motor functions of the elderly [13]. These vague recommendations clearly show that evidence on which to base age-specific design is lacking, and this lack of guidance leaves a significant gap in scientific knowledge. Because the human sensorimotor system changes naturally across the lifespan, such information is particularly crucial to the age-specific design of VEs.

 ISBN: 978-1-61208-250-9

B. Sensorimotor changes across the lifespan

The human body constantly changes throughout the lifespan. Most physiologic processes begin to decline at a rate of 1% per year beginning around age 30, and the sensorimotor system is no exception [14]. Both the processing of afferent information and the production of efferent signals steadily change as a function of age. Multiple studies demonstrate physical changes in brain tissues, in the excitability of the corticospinal tract and anterior horn cells, and in neurotransmitter systems [15-18]. A loss of neural substrate, including grey and white matter, occurs in both the cerebral cortex and the cerebellum [16, 17]. Tissue changes result in myriad functional changes within the central nervous system (CNS). A general deterioration of motor planning capabilities and feed-forward anticipatory control arises with aging [19-21]. Along with a decrease in planning ability, there also appears to be degradation of timing ability and general slowing of central processing [22, 23]. Loss of attentional resources also contributes to this slowing of central processing [24, 25]. The CNS re-weights sensory information when one source of feedback is compromised and compensates with alternative senses through a general systems neuroplasticity effect [26]. The implication here is important; the result of these attentional and processing changes is a decline in the ability to integrate multiple sensory modalities causing a relative decrease in the use of proprioceptive feedback and an increased reliance on vision for motor performance [25, 27, 28].

Comprehending the wide variety of physiologic changes occurring across the lifespan is important, but the concept of visual dominance for motor control in senior adults is especially imperative to the design of virtual environments. While VEs can recruit multiple senses for interaction, vision is the most common by far. Hence, provision of useful visual information for senior adults is extremely important to their success as a user group. While numerous studies characterize age-associated changes in visuomotor control, human movement is task specific. Thus, it is necessary to study human performance in VE surroundings [29, 30].

C. Aging and luminance contrast

Contrast sensitivity is one important aspect of visual function that declines significantly across the lifespan [31]. Additionally, it has been shown that reaction times to visual stimuli, a measure of general processing speed, vary based on the luminance contrast of those stimuli for both young and senior adults [32]. As reviewed in Section B, the decline of motor performance with aging results from the slowing of central sensorimotor processing. Since speed of processing varies with contrast level of the visual stimulus, it follows that motor performance will as well. Prior studies of the effects of luminance contrast on motor performance of young adults in natural environments support this concept [33, 34]. The current experiment seeks to extend these results by including a senior adult cohort, and by performing the experimental task within a virtual environment.

II. Methods

A. Participants

This research received approval from the University of Wisconsin-Madison Social and Behavioral Sciences Institutional Review Board under protocol number SE-2009-0112. Individuals participated in single experimental sessions totaling approximately 30 minutes each. We recruited the young cohort on the UW-Madison campus and the senior cohort on campus, throughout the general community, and from a local independent living apartment complex. Prior to participating in the study, individuals confirmed by self-report that they were healthy, living independently, without history of neurologic disease or injury, and right-handed. When a participant met all pre-screening criteria, s/he signed the informed consent form and filled out the General Practice Physical Activity Questionnaire, which provides an overall measure of functional level [35]. Participants also performed a standard visual acuity test using a Snellen chart [36]. Participants were required to demonstrate visual acuity (with glasses or contact lenses as needed) at a level of 20/40 or better. This represents the legal limit for driving in the State of Wisconsin. Participants then completed a modified version of the Simulator Sickness Questionnaire (SSQ) and the Mini-Cog [37, 38]. Participants were excluded if they experienced dizziness on a daily basis or showed evidence of dementia. Finally, we assessed computer via a seven point Likert scale ranging from "1-no prior experience" to "7-considerable experience." Thirteen participants met the criteria for the young adult group, and 13 for the senior adult group (Table I).

TABLE I. PARTICIPANT CHARACTERISTICS

Group	Mean age (yrs)	Age range	Male: Female	GPPAQ Activity Level
Young	21.2	19-24	5:8	Active
Senior	70.7	60-85	6:7	Moderately Active

Of the 13 senior participants sampled, three resided in the independent living apartment complex. No potential participants failed to meet the inclusion/exclusion criteria. Participants received an honorarium of $10 for completing the protocol.

B. Experimental apparatus

The Wisconsin Virtual Environment (WiscVE) was used to complete the experimental protocol (Fig. 1). The VE provides a head-coupled, stereoscopic experience to a single individual, allowing the user to grasp and manipulate augmented objects. A VisualEyez (PTI Phoenix, Inc.) motion analysis system, connected to a Windows PC workstation, captures three-dimensional motion information (e.g., movement of the subject's hand, head and physical objects within the environment) for real-time scene

 ISBN: 978-1-61208-250-9

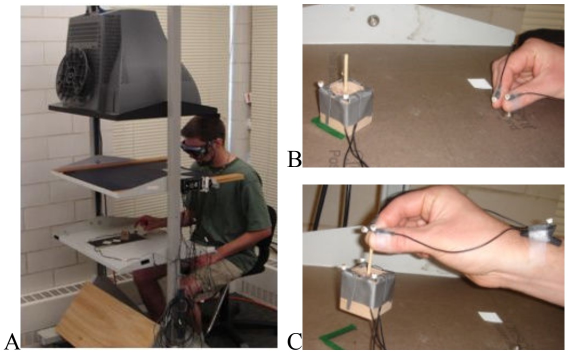


Figure 1. Panel A shows the WiscVE apparatus with downward facing monitor projecting to the mirror. Images are reflected up to the user wearing stereoscopic LCD shutter goggles, and thus the images appear at the level of the actual work surface below. Panels B and C demonstrate a reach to grasp task commonly utilized in this environment. The hand and physical cube are instrumented with light emitting diodes that are tracked by the VisualEyez (PTI Phoenix, Inc) system, not shown.

rendering and off-line kinematic analysis. The VisualEyez system monitors the 3D positions of small infrared light emitting diodes (LEDs) located on landmarks of interest. These landmarks include the tips of the thumb and index finger along with the radial styloid at the wrist to demarcate the hand. Objects in the environment are also equipped with three LEDs for motion tracking. Motion information from the VisualEyez system transmits on a subnetwork to a scene rendering Linux-based PC. Using the motion capture information, the scene is calculated and then rendered (10 ms lag time) on a downward facing CRT monitor placed parallel to a work surface. Also parallel to the computer monitor, a half-silvered mirror sits midway between the screen and the workspace to reflect images upward to the user. By wearing stereoscopic goggles, participants perceive the reflected image as if it were a three-dimensional object located in the workspace below the mirror**.** The environment is located in a dedicated room with blackout shades that provide the experimenter with good control of ambient room lighting.

C. *Procedure and design*

Participants completed a reach to grasp task, using a small wooden cylinder as the target object (3 mm diameter, 72 mm total height, Fig. 1B and1C). The visual scene varied from a low contrast condition equivalent to the individual's just noticeable difference contrast threshold, to the maximum contrast available in the WiscVE. Additionally, a moderate (50%) contrast condition added a third level to this factor (Fig. 2). To determine these settings, participants completed a JND contrast threshold test prior to the start of the experiment, as per the methods of Tamura, Satoh, Uchida, and Furuhata [39]. Briefly, participants sat for five minutes at the environment console to allow for dark adaptation. At this point, we instructed participants to begin increasing the luminance of the target object and finger representation gradually by tapping their index finger on a virtual button. Individuals performed this task in a self-

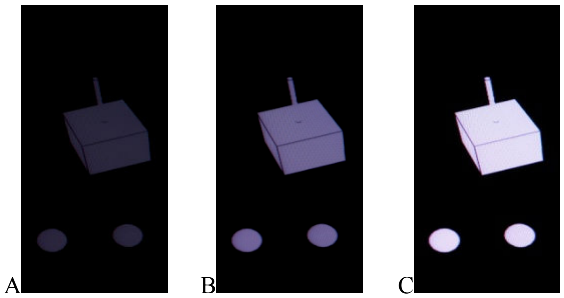


Figure 2. The visual scene varied in luminance contrast from A) the just noticeable difference threshold, to B) 50% contrast level, to C) 100% maximum luminance contrast.

paced manner until they just perceived the presence of the object and their fingertips in the environment. Each participant performed three JND trials, and the average luminance value at which they perceived the target and finger spheres determined their individual threshold. Luminance was controlled using the standard 255-point Red-Green-Blue (RGB) scale, maintaining equal values for each of the three color values, thereby maintaining white (achromatic) visual stimuli. The moderate contrast level was calculated by taking the difference between the maximum RGB (i.e., 255) and the individual's JND threshold, dividing this by 2, and adding the result to the JND value. For example, a participant that, on average, perceived the target and fingers when they reached an RGB level of 21 would have a moderate contrast condition set at: ((255-21)/2) + 21 = 138. This normalizes the moderate level to the individual. Actual luminance values in candela per square meter (cd/m^{2}) for the background and stimuli were obtained using a spot luminance photometer (Spectra Cine). JND values ranged from twelve to 26 on the RGB scale. RGB and luminance values are listed in Table II.

TABLE II. VISUAL STIMULUS PARAMETERS

RGB (JND-Low)	c/m^2	RGB (Mod)	cd/m^2
12	0.02	133.5	23.24
13	0.03	134	23.49
14	0.04	134.5	23.74
15	0.05	135	24.00
16	0.05	135.5	24.25
17	0.06	136	24.51
18	0.08	136.5	24.77
19	0.09	137	25.03
20	0.10	137.5	25.29
21	0.12	138	25.55
22	0.14	138.5	25.81
23	0.15	139	26.08
24	0.17	139.5	26.35
25	0.20	140	26.62
26	0.22	140.5	26.89
RGB (High)	cd/m^2		
255	147.34		

The task instructions for this movement were for participants to reach to grasp the target object with a precision grasp (i.e., thumb and index finger), initiating the movement as quickly as possible after appearance of the target. The reach to grasp progressed in a mid-sagittal plane, in line with the middle of the right shoulder. Subjects started from a designated start mark, identified with a tactile cue (a small metal cap nut, 8 mm in diameter, 7 mm in height). The target object rested 180 mm away from the start mark, resulting in a visual eccentricity of approximately 10 degrees from midline. Each trial began with the participant resting their hand away from the start mark. A visual cue, in the form of a green circle in the far right corner of the desktop, indicated when to move to the start position. When the participant placed their hand in the correct position over the start mark, with their thumb and index finger resting together in a light pinch grip, the visual cue disappeared. The system requires participants to maintain this position for one second prior to the start of a trial, to ensure individuals start from complete rest. Once the one-second rest period had elapsed, the system initiated the trial. To prevent anticipation, the presentation of the target object and fingers varied between 400-800 ms after trial initiation. Participants then reached out to grasp the cylinder, lifting it vertically about 50 mm before replacing it on the desktop. After replacing the object, participants moved their hand near the start mark, awaiting the visual cue for initiation of the next trial.

Participants completed three practice trials in each condition prior to the start of the experiment. There were ten trials per contrast condition, with a 2 (age group) x 3 (contrast level) design, for a total of 30 experimental trials presented in random order. These reaction time trials were embedded in a series of other experiments reported elsewhere.

D. Dependent measures

The primary dependent variable of interest was simple reaction time (RT). Subtracting the time point of stimulus presentation from the time point of movement onset provided the RT results. Movement onset was defined as the point where the fingers had departed from their starting position by 5 mm. This corresponds to the amount of displacement needed to depress a key on a standard computer keyboard, which is commonly employed in studies of RT.

E. Hypotheses

Based on prior studies of luminance contrast and reaction times, as well as contrast sensitivity and age, the following was hypothesized:

Hypothesis 1: Young adults will have faster reaction times than senior adults.

Hypothesis 2: Both young and senior adults will improve their reaction times as luminance contrast increases.

Hypothesis 3: Young adults will improve their reaction time at the moderate and high contrast levels, seniors only at the highest contrast.

F. Data analysis

SPSS (IBM) was used for statistical analysis. We performed a Mann-Whitney U test to analyze group difference in computer experience. Next, an ANOVA with post-hoc blocking was used to assess group difference in RT controlling for group difference in computer experience. Computer experience (3 levels based on Likert responses: 7, 6, and ≤ 5) was used as the blocking factor. This allowed evaluation of hypothesis 1. Planned comparisons of contrast level within each age group evaluated hypothesis 2 and 3. For all results, the a priori alpha level was set at $p \leq 0.05$.

III. RESULTS

Computer experience level was significantly different between groups (Young Median = 7, Mode = 7, Senior Median = 6, Mode = 6, $p = 0.04$). There was a significant main effect of age ($F_{1, 20} = 4.55$, $p = 0.05$) controlling for computer experience, with young adults having faster reaction times (M = 441 ms, SD = 90.92) than senior adults (M = 498 ms, SD = 90.45). The simple effect of contrast within each age group appears in Fig. 3. This shows that for young adults, the low contrast reaction time was significantly slower than both the moderate and high contrast conditions, which did not differ. For senior adults, the low and moderate contrast conditions did not significantly differ, while both the low to high and moderate to high mean differences did reach statistical significance.

IV. DISCUSSION

The first important finding is that young adults have faster reaction times than senior adults, replicating numerous previous studies and confirming hypothesis 1 [22, 32]. Further, level of computer experience differed between groups, but age group differences remained apparent in the analysis. This lends support to the idea that group differences are truly related to sensorimotor differences between groups, rather than familiarity with technology.

The data clearly show that reaction times improve for both groups with increasing luminance contrast (Fig. 3). While young adults improve from low to moderate contrast, they do not get further benefit going from moderate to high contrast. Senior adults, on the contrary, improve significantly in the high contrast condition, but do not improve going from low to moderate. This confirms hypothesis 2 and 3.

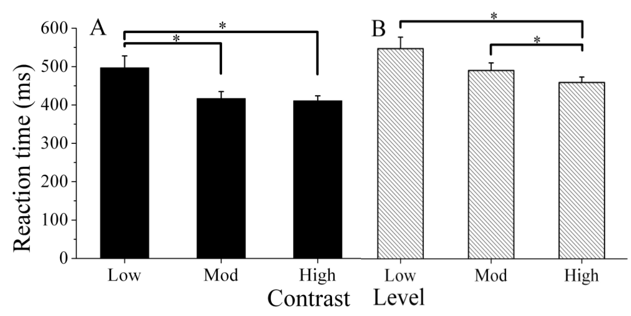


Figure 3. The simple effect of contrast level on reaction times for **(A)** Young adults and **(B)** Senior adults.

Previous study of reaction times, contrast, and age demonstrated no age group difference in the pattern of contrast-mediated reaction time improvement [32]. The results in the current study differ from that finding, and are more consistent with the concept that contrast sensitivity changes across the lifespan [31]. Accordingly, luminance contrast must be accounted for when designing interfaces for specific user groups.

Considering reaction time as a measure of general sensorimotor processing speed, and the strong link between motor performance and processing speed, these results support the idea that improvements in motor performance may be realized through the age-specific utilization of luminance contrast levels. It is important to note that while young adults achieved their entire performance enhancement with only moderate contrast, seniors required the extra increase to the high contrast condition to maximize their gains (Fig. 3). This pattern indicates that young adults can make better use of lower contrast visual stimuli, having a performance ceiling somewhere below the 50% contrast level, while senior adults may not have a distinct ceiling within the ranges of luminance contrast available in the WiscVE. This concept will require further exploration of other VE tasks under varying conditions of luminance contrast with more complex kinematic analyses.

Given the results of the current study, it may be tempting to conclude that the highest available contrast should always be used for motor tasks performed in virtual environments. While further study is needed, it must be recognized that high contrast comes at a cost. First, high luminance contrast, by definition, requires that at least one portion of the visual scene be presented at maximum luminance. This could potentially lead to eyestrain, limiting the timeframe for comfortable use of the VE. This may be especially important for users that are hypersensitive to light. Second, such contrast restrictions limit the color combinations available for use. Strictly speaking, the highest available contrast is black and white. While this may be acceptable in certain circumstances, it is not likely to be common. This is especially apparent in the design of VEs for use with young adult populations. To limit this age group to high contrast conditions would unnecessarily limit the potential richness of the visual display, adding no extra motor control benefit over a moderate contrast visual scene.

Limitations

A limitation exists with the assessment of visual acuity in this study. The chart used provides a standard measure of acuity at a distance of 20 feet, while the experimental task occurs within the personal space of the participant. A number of other sophisticated tests of vision exist, gauging such functions as depth perception and figure-ground discrimination, and may provide valuable information worth considering in future studies. Next, the method of defining each individual’s just noticeable difference contrast threshold is inherently subjective. While carefully scripted instructions were used to describe the procedure, each individual had to interpret the meaning of “just able to perceive,” and this introduces a source of error in defining contrast levels. Additionally, the low and moderate contrast levels were set relative to each individual’s personal contrast threshold, while the high contrast condition was set to the limit of the environment. This means that the high contrast condition was relatively different for each subject. However, given the narrow range of JND results, the relative differences are quite small. Finally, although every reasonable attempt was made to block ambient light from windows in the Human Motor Behavior lab, some minimal fluctuation did occur. This did not result in any cases of participants reporting a change in their ability to perceive the visual scene.

V. CONCLUSION AND FUTURE WORK

The virtual environment design process must account for the age of the targeted user when considering parameters of visual scene rendering. Luminance contrast is one property that is easily programmable, and has a clear age-specific effect on motor performance. Younger adults have a wider bandwidth of contrast levels that may result in optimum sensorimotor performance. Older adults, on the contrary, need higher levels of contrast to experience their maximum motor performance benefits. The exact range is yet to be determined for either group, but will be the subject of future research. General VE implications include the use of high luminance contrast to improve sensorimotor processing speed. This is an important consideration when targeting user groups known to be deficient in speed of processing, such as senior adults. Additionally, increased processing speed with high contrast stimuli is likely to improve performance when tasks are of particularly high complexity or difficulty. This will be further investigated in future research as well. Potential consequences of high contrast visual stimuli, such as eyestrain, were not assessed in this experiment. This also warrants further user-centered study, and should be carefully considered along with both the intended application and the end-user group of any high contrast virtual environment display.

ACKNOWLEDGMENT

The authors thank Mr. Drew Rutherford for his valuable assistance with experimental design and data collection. Thanks to Stephanie Ehle and Nicole Wenman for data collection and analysis assistance. This work was funded by the National Science Foundation, grant number 0916119.

REFERENCES

[1] P. A. Heidenreich, J. G. Trogdon, O. A. Khavjou, et al., "Forecasting the Future of Cardiovascular Disease in the United States A Policy Statement From the American Heart Association," Circulation, vol. 123, 2011, pp. 933-944.

[2] C. J. Bohil, B. Alicea, and F. A. Biocca, "Virtual reality in neuroscience research and therapy," Nature Reviews Neuroscience, 2011.

[3] A. S. Rizzo and G. J. Kim, "A SWOT analysis of the field of virtual reality rehabilitation and therapy," Presence: Teleoperators & Virtual Environments, vol. 14, 2005, pp. 119-146.

[4] N. E. Seymour, A. G. Gallagher, S. A. Roman, et al., "Virtual reality training improves operating room performance: results of a randomized, double-blinded study," Ann. Surg., vol. 236, 2002, pp. 458.

[5] R. T. Hays, J. W. Jacobs, C. Prince, and E. Salas, "Flight simulator training effectiveness: A meta-analysis," Mil. Psychol., vol. 4, 1992, pp. 63-74.

[6] M. K. Holden, "Virtual environments for motor rehabilitation: review," Cyberpsychol Behav., vol. 8, 2005, pp. 187-211.

[7] L. Liu, B. Watson, and M. Miyazaki, "VR for the elderly: quantitative and qualitative differences in performance with a driving simulator," CyberPsychology & Behavior, vol. 2, 1999, pp. 567-576.

[8] D. L. Jaffe, "Use of virtual reality techniques to train elderly people to step over obstacles," in Technology and Persons with Disabilities Conference, Los Angeles, CA, 1998.

[9] K. M. Stanney, R. R. Mourant, and R. S. Kennedy, "Human factors issues in virtual environments: A review of the literature," Presence, Teleoperators & Virtual Environments, vol. 7, 1998, pp. 327-351.

[10] N.I. Durlach and A.S. Mavor, "Virtual reality: Scientific and technical challenges." National Academy Press, Washington, D.C., 1995.

[11] P. J. Grabowski and A. H. Mason, "Sensory feedback for motor performance in virtual environments across the lifespan," in Proceedings of the IADIS International Multiconference on Computer Science and Information Systems, Rome, Italy, 2011.

[12] C. Allen R, J. Singer M, P. McDonald D, and E. Cotton J, "Age Differences in a Virtual Reality Entertainment Environment: a Field Study," Human Factors and Ergonomics Society Annual Meeting Proceedings, vol. 44, 2000, pp. 542-545.

[13] W. Karwowski (Ed.), International Encyclopedia of Ergonomics and Human Factors. Boca Raton, FL: CRC Press, 2006.

[14] L. J. Schut, "Motor system changes in the aging brain: what is normal and what is not," Geriatrics, vol. 53 Suppl 1, 1998, pp. S16-9.

[15] H. K. Kuo and L. A. Lipsitz, "Cerebral white matter changes and geriatric syndromes: is there a link?" J. Gerontol. A Biol. Sci. Med. Sci., vol. 59, 2004, pp. 818-826.

[16] B. B. Andersen, H. J. Gundersen, and B. Pakkenberg, "Aging of the human cerebellum: a stereological study," J. Comp. Neurol., vol. 466, Nov 17, 2003, pp. 356-365.

[17] N. Raz and K. M. Rodrigue, "Differential aging of the brain: Patterns, cognitive correlates and modifiers," Neurosci. Biobehav. Rev., vol. 30, 2006, pp. 730-748.

[18] P. Rossini, M. Desiato, and M. Caramia, "Age-related changes of motor evoked potentials in healthy humans: non-invasive evaluation of central and peripheral motor tracts excitability and conductivity," Brain Res., vol. 593, 1992, pp. 14-19.

[19] J. H. Yan, J. R. Thomas, and G. E. Stelmach, "Aging and rapid aiming arm movement control," Exp. Aging Res., vol. 24, 1998, pp. 155-168.

[20] A. Sterr and P. Dean, "Neural correlates of movement preparation in healthy ageing," Eur. J. Neurosci., vol. 27, 2008, pp. 254-260.

[21] J. H. Hwang, Y. T. Lee, D. S. Park, and T. K. Kwon, "Age affects the latency of the erector spinae response to sudden loading," Clin. Biomech, vol. 23, 2008, pp. 23-29.

[22] T. A. Salthouse, "Aging and measures of processing speed," Biol. Psychol., vol. 54, 2000, pp. 35-54.

[23] N. Wild-Wall, R. Willemssen, M. Falkenstein, and C. Beste, "Time estimation in healthy ageing and neurodegenerative basal ganglia disorders," Neurosci. Lett., vol. 442, 2008, pp. 34-38.

[24] A. Kluger, J. G. Gianutsos, J. Golomb, et al., "Patterns of motor impairement in normal aging, mild cognitive decline, and early Alzheimer's disease," J. Gerontol. B Psychol. Sci. Soc. Sci., vol. 52, 1997, pp. P28-39.

[25] D. J. Goble, J. P. Coxon, N. Wenderoth, A. Van Impe, and S. P. Swinnen, "Proprioceptive sensibility in the elderly: Degeneration, functional consequences and plastic-adaptive processes," Neurosci. Biobehav. Rev. 2008.

[26] S. Heuninckx, N. Wenderoth, and S. P. Swinnen, "Systems neuroplasticity in the aging brain: recruiting additional neural resources for successful motor performance in elderly persons," J. Neurosci., vol. 28 2008, pp. 91-99.

[27] S. Chaput and L. Proteau, "Aging and motor control," J. Gerontol. B Psychol. Sci. Soc. Sci., vol. 51, 1996, pp. P346-55.

[28] R. O. Coats and J. P. Wann, "The reliance on visual feedback control by older adults is highlighted in tasks requiring precise endpoint placement and precision grip," Experimental Brain Research, vol. 214, 2011, pp. 139-150.

[29] L. Proteau, "On the specificity of learning and the role of visual information for movement control." in Vision and Motor Control, Proteau, L. and Elliott, D (Eds.), Amsterdam: Elsevier, 1992, pp. 67-103.

[30] M. Mon-Williams and G. P. Bingham, "Discovering affordances that determine the spatial structure of reach-to-grasp movements," Experimental Brain Research, 2011, pp. 1-16.

[31] J. Faubert, "Visual perception and aging," Canadian Journal of Experimental Psychology, vol. 56, 2002, pp. 164-176.

[32] V. Porciatti, A. Fiorentini, M. C. Morrone, and D. C. Burr, "The effects of ageing on reaction times to motion onset," Vision Res., vol. 39, 1999, pp. 2157-2164.

[33] M. M. Veerman, E. Brenner, and J. B. J. Smeets, "The latency for correcting a movement depends on the visual attribute that defines the target," Experimental Brain Research, vol. 187, 2008, pp. 219-228.

[34] M. Spering, D. Kerzel, D. I. Braun, M. J. Hawken, and K. R. Gegenfurtner, "Effects of contrast on smooth pursuit eye movements," Journal of Vision, vol. 5, 2005, pp. 455-465.

[35] Physical Activity Policy and Health Improvement Directorate, "The general practice physical activity questionnaire (GPPAQ): a screening tool to assess adult physical activity levels, within primary care," 2009.

[36] H. Snellen, "Probebuchstaben zur Bestimmung der Sehschärfe," Utrecht, P. W. van de Weijer, 1862.

[37] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," The International Journal of Aviation Psychology, vol. 3, 1993, pp. 203-220.

[38] S. Borson, J. Scanlan, M. Brush, P. Vitaliano, and A. Dokmak, "The Mini-Cog: a cognitive'vital signs' measure for dementia screening in multi-lingual elderly," Int. J. Geriatr. Psychiatry, vol. 15, 2000, pp. 1021-1027.

[39] T. Tamura, T. Satoh, T. Uchida, and T. Furuhata, "Quantitative evaluation of luminance nonuniformity mura in LCDs based on just noticeable difference (JND) contrast at various background luminances." IEICE Transactions on Electronics, 89(10), 2006, pp. 1435-1440.

Networked Visibility: The case of smart card ticket information

Maja van der Velden, Alma Culén, Jo Herstad and Atif Abdelhakeem
Department of Informatics, Group for Design of Information Systems
University of Oslo
Oslo, Norway
majava@ifi.uio.no, almira@ifi.uio.no, johe@uio.no, abdelhat@student.matnat.uio.no

***Abstract*—This paper concerns the replacement of paper tickets with smart card tickets for public transportation. By contrasting the visibility of ticket information to users of paper tickets and smart card tickets, this paper describes the move from local information on paper tickets to distributed information on smart cards. Using the concept of 'networked visibility', this paper argues that this move has resulted in less informed travelers and more informed providers. In order to restore the accessibility of ticket information to users, we present one possible solution, ARTick, a mobile phone smart card reader app. ARTick shows that smart cards can make complete ticket information visible to the user, whenever and wherever this information is needed.**

Keywords—Information Visibility; Networked Visibility, Smart Cards; Ticket Information; Mobile Phones

I. INTRODUCTION

A major research and development area for human-computer interaction (HCI) over the last years has been to explore ways of interaction that emerge by embedding computing and networking technology in the everyday physical world in which we live. This development goes by many names such as tangible interaction [6], ubiquitous computing [5] and embodied interaction [3]. Examples of this development are pen and paper with embedded computing technologies and transportation tickets with embedded computer chips with networking capabilities.

When computing and network technologies are embedded in everyday objects, there are many ways *"to capitalize on our familiarity, skill and experience in dealing with the everyday world around us"* writes Dourish [2]. New values, new possibilities, but also new concerns may emerge from interaction between these familiar objects with and without embedded technology.

This paper reports from a case in which paper-based tickets for public transportation are replaced by so-called *contactless smart cards*. A smart card is a credit card size plastic card containing a microchip with antenna for contactless communication with a card reader; see Fig. 1 for the card and Fig. 2 for readers. The computerized and networked smart card can hold different kinds of information: ticket information, monetary information (the amount of money put on the card), but also other information such as the ticket for a football match or the annual subscription to a museum.

Our focus is on the visibility of information that is needed to use the smart card as a valid transportation ticket.

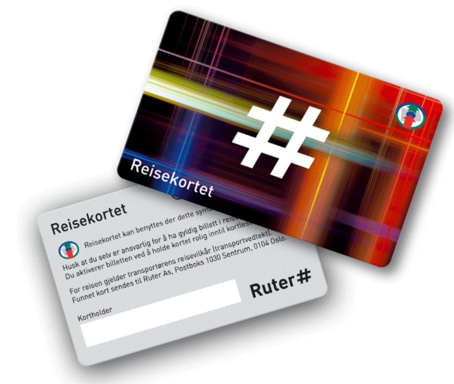


Figure 1. A smart card used in public transportation in Norway.

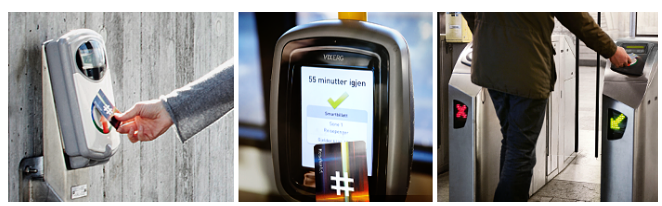


Figure 2. Smart card readers.

We identified three basic ticket information needs of public transportation users: the type of the ticket, the value of the ticket, and the duration of the ticket. There are other ticket information needs, such as the price of one trip, an overview or log of implemented trips, an overview of past travel expenses, etc., but they don't add significantly to our argument.

In design of information systems and usability studies, the visibility of information is related to the visibility of a system's status. Being informed about a system's status is one of the ways in which users receive feedback on a system's use or performance. Studies of the visibility of information on smart cards have been implemented in several sectors, such as supply chain management, the automotive industry, and the healthcare sector [1, 10, 11, 12]. We are not aware of visibility studies of ticketing systems in the public transportation sector although some other interesting and related issues have been reported in [7, 8, 9]. Morgner *et al.'s* proposal [9] is of particular interest to us, as the authors propose similar technologies in their design solution.

The rest of this paper is organized as follows: in the next section we present our case, the move from a paper-based ticket to a computerized ticket. In Section III we report on a small explorative study among randomly selected inhabitants of Oslo, conducted in order to get an indication if the visibility of smart card information influences users experience with the new system. In Section IV, we will use the concept of networked visibility, first introduced by Stalder [14], to discuss horizontal and vertical visibility of ticket information. In Section V we present ARTick as one possible solution for the visualization of smart card-based ticket information. In the last section, we present our conclusions and future work.

II. From Paper Tickets to Smart Card Tickets

Moving from paper-based practices [13] or desktop-based practices to practices where computing and communications technologies are embedded into everyday objects poses many challenges. Dourish [2] states: "*In this world, our primary experience of computation is not with a traditional desktop computer, but rather with a range of computationally-enhanced devices, pieces of paper, pens, walls, books, hammers etc*." Rust and Kannan [12] consider this ubiquitous computation to be a fantastic opportunity to enhance customer experiences. In the case of smart cards, all sort of information may be stored and used by service provider(s) in order to offer better one-to-one services. We discuss how such embedded technology often offers much wider opportunities to providers than to customers.

A. The visibility of information

In our study of the smart card we focus on the visibility of public transportation ticket information. Our case is the 'Reisekortet', the smart card ticket used in the public transportation system in Oslo, Norway. The system was first introduced in 2009, replacing the paper ticketing entirely.

We are interested in the following questions: How are users of paper tickets able to find answers to their three basic ticket information needs: type of the ticket, value of the ticket and duration of the ticket (see Table 1) and how has this changed with the introduction of the smart card ticket?

We will answer this question by looking into three actions the users engage in: purchasing the new ticket, using a valid ticket and having an expired ticket.

B. Purchasing a ticket

When a ticket is purchased, the three pieces of basic information (type of the ticket, value of the ticket and duration of the ticket) are given by the user to a sales person or are selected by the user on a vending machine or on the public transportation website. In addition, public transportation users have also the choice between a registered and an unregistered ticket. 'Registered' means that the name and date of birth of the user is registered with the public transportation provider. 'Unregistered' means that the user is anonymous and that the age of the user is unknown. All information given/selected by the user is registered on the ticket.

TABLE I. Basic ticket information needs users of public transportation have

Type of ticket The type of the ticket refers to the different kinds of tickets available. We can differentiate between types of tickets based on the number of trips and types of tickets based on the particular period they cover independent of the number of trips (day, week, month, and year). Other types are registered or unregistered (anonymous) tickets, and regular and discount tickets. • A popular paper ticket was the unregistered 8 trip-ticket, the so-called *flexi card*, which was available as a regular ticket and a discount ticket. The flexi card could be used by more than one traveler at the same time. There is no smart card variation of this ticket. • A popular smart card ticket is the prepaid card, which can be topped up when needed. This ticket can only be used by one traveler at the time. There is no paper variation of this ticket.
Value of ticket The type of ticket decides the monetary value of the ticket. In the case of the pre-paid fill-up ticket, the value of the ticket depends on how much money the user has put on the ticket. The monetary value of all tickets diminishes with use. A ticket has zero value when the duration of the ticket has expired or when the monetary value is below the price of a ticket. In the case of prepaid cards, any amount less than the value of a single ticket may be left on the card. Paper tickets did not have this characteristic.
Duration of ticket The duration of the ticket is decided by the date and time stamp of a ticket and varies for the different ticket types. Registered monthly paper tickets were sent automatically by mail to the user before the monthly ticket expired. Registered smart cards can be automatically topped up (in case of a prepaid card) or extended (in case of the 30 days card).

All this information is at all times visible on the paper ticket in the form of printed text (type, value, duration), the size of the ticket (type), the color of the ticket (type), and the shape of the ticket (type). For example, the 8-trip ticket (see Fig. 3) was the only folded paper ticket. It had a pre-printed text to indicate the value of the ticket (kr.180) and the word ‘voksne’ (adults) to indicate that it was a regular ticket. The fact that it was a regular ticket was also indicated by its color combination.

The printed text on a strip is a timestamp, indicating the time when the one-hour validity of the ticket ends. This timestamp is added to the ticket when a traveler enters a metro platform or a bus or tram and inserts the card in a ticket stamp machine (see Fig. 3).

Figure 3. A paper ticket (left) and a ticket stamp machine (right).

The crucial difference between the paper ticket and the smart card ticket is that on the smart card ticket, information is distributed across several devices and places, but it is never visible on the ticket itself. The information can *become* visible in four different ways: via stationary ticket readers positioned at the entrances of stations and platforms of the metro and inside busses and trams, scanners handheld by human ticket controllers, smart card terminals at the point of purchase, and the Internet (only for registered smart card holders, Fig. 4).

C. *Using a Ticket*

When one is travelling, the value and duration of the ticket change. On the paper ticket this information is at all times visible, while travelers with a smart card need to use ticket readers to access this information on their card. The stationary readers are also used to validate a ticket and give information about the type of card, expiration date or remaining value of the card, and expiration time. This information is visible for two seconds at the time of validation. This is often too short. The user can wait 2 minutes after validation to display the information again. The fact that the type of ticket is not visible without scanning it, presents the risk of traveling with a wrong card, e.g. a parent can use a child ticket without knowing it.

When a traveler validates a ticket, the reader can provide the wrong information. For example, an 11-year-old girl, who travels alone on a tram to her dance school, uses her prepaid smart card twice a week. Incidentally, her mother accompanies her one-day and notices that the child pays the adult fee instead of the discounted fee for children. The child’s birthday was recorded at the time of the purchase of the smart card ticket and the card has been working well over a long period of time. The mother and the daughter walk into the public transportation service centre. The customer representative scans the card. All the trips, and the fees paid for them, appear on the screen. It becomes apparent that somehow the discount child’s smart card was read as a regular card. The customer representative counts the number of wrongly charged trips, fills a paper based refund form, and issues the overcharged amount of fees in cash.

D. *The Expired Ticket*

A paper ticket is expired when the timestamp on a ticket has expired. The user of a smart card will not be able to see if the ticket is still valid. The ticket has to be read (see above). If the ticket is a registered smart card, the validity of the ticket can also be checked by logging onto the public transportation system’s website.

At the moment, travelers have no way of checking the validity of their smart card ticket when they leave their home or office unless they have a registered card and Internet access.

Figure 4. Accessing on the internet the information from the smart card.

Smart card users taking the bus or tram find out if their ticket has expired or not by using a card reading located inside the bus or tram. Our observations with smart card readers located with the bus driver made clear that many travelers are surprised to find out that they have not enough funds on their card and that they were attempting to travel with an expired card (see Fig. 5).

In those cases the travelers need to buy an expensive one-time paper ticket from the bus driver or they have to leave the bus.

Figure 5. Validating a smart card ticket on the bus.

III. WHAT SMART CARD TICKET USERS SAY

As users of the public transportation system in Oslo, the authors are familiar with the change in the visibility of ticket information since the introduction of the ticket smart card. We decided to ask other users what their experiences with the new smart card ticket were. We selected at random 20 people in Oslo to conduct quick, semi-structured interviews about their awareness of the information stored on their smart card ticket. Eight of the twenty people interviewed were interviewed at the public places, but away from transportation points, in order to get a feeling if they would remember how much money they had or when they used their smart ticket last. Three persons were monthly ticket users and knew the expiration date of their card precisely. They also knew when they used the transportation last, though this piece of information was not important for them. Two were using a prepaid card, which can be used for multiple trips as long as there are enough funds on it. They also felt that they had no problems keeping track of the amount of money left on their respective cards. A young student, using the same type of prepaid card said: "*I never know, getting on the tram, if I actually have any money on the ticket. It is always a game of chance"*. One person in this group was an out of town visitor who always buys a day ticket because *"everything else is too complicated for me."* She continued to explain that it is hard to keep track of when she last used the transportation (tickets have 1 hour validity). As she was not in Oslo very often, she always had multiple tasks at different points of the city to accomplish. The day ticket then was the best option, also because the ticket could then be discarded, instead of taken care of for the future use. The last person we approached said that he has never used public transportation in Oslo.

The remaining twelve persons were interviewed at a bus stop (four) and at a metro station (eight). Ten of them are monthly ticket users. Three persons from this group said that they remember with certainty when their ticket expires. One of them remarked that although she does remember, she wishes that she did not have to. Three said that they remember approximate expiration date. They scan the card when it comes close to the day they think the ticket will expire. Three said that they do not remember expiration date and have to check their ticket frequently. One older person said that she keeps the receipt from the sale of the card in her valet and looks at it occasionally, as the date of the purchase is printed on it. Two persons with prepaid cards said that they have problems remembering how much money they still have on their card. One of them in particular had difficulties as he is only an occasional user of the public transportation system and forgets how much money is left on the card between the trips.

Thus, only about half of the people we talked to were entirely comfortable with the visibility aspect of the smart ticketing system. This number increases if those who do not mind scanning their card occasionally are included in the group that feels comfortable. However, 1/5 of this sample felt clearly uncomfortable with smart card information visibility.

Although only 20 persons were interviewed, the results give an indication that information visibility of a smart card is a real issue. The user satisfaction with the smart card ticketing solution may be improved by offering a better visibility of ticketing information.

IV. THE NETWORKED VISIBILITY OF TICKET INFORMATION

The answer to a simple question "*how much money is left on my smart card?"* involves a variety of devices and places. Ticket information, once located on a piece of paper in the hand of the user, is now distributed and networked. Stalder [14] calls this *networked visibility*, which is "created by the capacity to record, store, transmit, access communication, action, and states generated through digital networks". Stalder studies Web 2.0 and presents two types of visibility: *horizontal visibility* pertaining to information becoming visible to users and *vertical visibility* pertaining to what information the service providers can see. While users can manage their horizontal visibility, i.e. what information about themselves becomes visible to others, they have no control over the vertical visibility of their information. Service providers have access to the information of all users, but this visibility is one way, it is invisible to the users.

Similar to Stalder, we can differentiate between the horizontal and vertical visibility of ticket information. We understand *horizontal visibility* as the visibility of ticket information to the user of the public transportation system. The paper ticket user has immediate horizontal visibility, at all times and places. The ticket information is directly visible on the paper ticket – when the ticket is in use, not in use, or expired. The smart card user's horizontal visibility is limited to particular places: when the ticket is purchased, when it is read or scanned, or when it is checked on the web (only for registered cards). As we saw in the previous section, many travelers are insecure about the status of their smart card: they are not sure if the card is (still) valid.

We understand *vertical visibility* as the visibility of the ticket information to the provider of the public transportation system. The provider has other ticket information needs than the user. The provider is interested in *use information*, such as the users' frequency, time, and destination of travel, and what type of ticket they use. This information is the basis for organizing public transportation schedules, the frequency of departures, and the number of routes. The provider had only limited vertical visibility when the paper tickets were in use and therefore had to implement user surveys to get this information. With the introduction of the smart card, the provider has full access to ticket information.

In our case, the networked visibility created by embedding computing and networking capabilities to a transportation ticket has decreased the horizontal visibility of the users and significantly increased the vertical visibility of the provider. The loss of horizontal visibility negatively affects a large number of travelers who are uncomfortable using their smart card. They can't transfer their familiarity, skill, and experience in using the paper ticket to the smart card.

The increase of the vertical visibility of the provider creates new concerns in terms of privacy as the whereabouts of registered smart card users is recorded, stored, and transmitted. These data can be accessed or aggregated for uses not directly related to the public transportation system. As we saw above, an employee at a service point could access such data. On the other hand, the availability of these data made it possible for the mother to get a refund. This would have been impossible with a paper ticket.

V. ARTick: Augmenting the Smart Ticket

The design of the smart card ticket builds forth on some of the characteristics of the paper ticket: tickets need to be validated before use and the points of validation are at the same locations as the paper ticket. The main difference is the visibility of ticket information. On paper tickets, information was visible at all times and places because it was *local information*, it was locally stored on the paper ticket. Can we make this characteristic of the paper ticket available on the smart card ticket?

In order to improve the user experience with smart tickets and offer local visibility of the ticketing information, we made a simple prototype: ARTick (Augmented Reality Ticket). ARTick turns any smart phone into a mobile smart card reader using NFC (Near Field Communication) standards, see also [9]. NFC is a short-range wireless technology, enabling one-way and two-way communication between smart phones or between smart phones and other wireless devices, in our case the contactless smart card [4].

On NFC-enabled mobile phones, the ARTick application uses NFC to read the information off the card. The application enables the user to check the type of ticket, the value and duration of the ticket, as well as the latest transactions. ARTick enables ticket information to be read in 2D and 3D, augmented using the camera as shown in Fig. 6, as well as audio for the visual impaired.

Non-NFC-enabled smart phones use the camera to take an image of the card number on the back of the ticket smart card and use Optical Character Recognition (OCR). This card number corresponds with ticket information stored on the website of the public transportation provider. The same ticket information will now be available on NFC-enabled smart phone. The addition of audio is in particular interesting for users who have issues with their vision, whether it is related to sight challenges or various forms of dyslexia. ARTick follows universal design principles [15]. ARTick enables the user to leave home with a valid smart card ticket; this may result in more confident and informed users of the public transportation system.

VI. Conclusion and Future Work

New concerns emerged after the move from paper tickets to smart card tickets. Users couldn't use their familiarity with the paper ticket in their use of the smart card ticket, because this familiarity was based on their immediate and continuous access to ticket information, type of ticket, value of ticket, and duration. The smart card ticket didn't provide this type of access to the user. At the same time, the provider gained access to user and use information, creating new concerns about privacy.

We have used the concepts of networked visibility and horizontal and vertical visibility to discuss the emergence of new concerns when computing and networking technology is embedded in a transportation ticket. These concepts were also used to explore solutions, such as how to restore the horizontal visibility of ticket information, the immediate and continuous ticket information to the users. ARTick, an app that turns a smart phone into a smart card reader, provides one possible solution to the problem.

As to future work, we plan to study the effect of ARTick on user satisfaction with a smart card transportation system. Further, we wish to use the concept of networked visibility to study information accessibility in other systems using solutions similar to the smart card.

References

[1] B. A. Aubert and G. Hamel, "Adoption of smart cards in the medical sector:: the Canadian experience," *Social Science & Medicine*, 53, (7), (2001), 879–894.

[2] P. Dourish, What we talk about when we talk about context, *Personal Ubiquitous Computing, 8*, 1 (2004), 19–30.

[3] P. Dourish, *Where the action is; the foundations of embodied interaction*. MIT press, Boston, MA, USA, 2001.

[4] D. K. Finkenzeller, RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication, Third Edition. John Wiley & Sons, 2010.

[5] A. Greenfield, *Everyware: The Dawning Age of Ubiquitous Computing*. Peachpit Press, Berkeley, CA, USA, 2006.

[6] E. Hornecker. and J. Buur, "Getting a grip on tangible interaction: a framework on physical space and social interaction". Proceedings of the SIGCHI conference on Human Factors in computing systems, CHI '06, ACM Press (2006), 437–446.

[7] H. Iseki, A. Demisch, B. D. Taylor, and A. C. Yoh, "Evaluating the Costs and Benefits of Transit Smart Cards," *PATH Research Report*, Aug. 2008.

[8] N. Mallat, M. Rossi, V. K. Tuunainen, and A. Öörni, "An empirical investigation of mobile ticketing service adoption in public transportation," *Personal and Ubiquitous Computing*, vol. 12, no. 1, pp. 57–65, Nov. 2006.

[9] F. Morgner, D. Oepen, W. Muller, and J. P. Redlich, Mobile smart card reader using NFC-enabled smartphones" in A. U. Schmidt, G. Russello, I. Krontiris, and S. Lian, Eds., Security and Privacy in Mobile Information and Communication Systems, Springer 2012.

[10] C. R. Plouffe, J. S. Hulland, and M. Vandenbosch, "Research Report: Richness Versus Parsimony in Modeling Technology Adoption Decisions--Understanding Merchant Adoption of a Smart Card-Based Payment System," *Information Systems Research*, vol. 12, no. 2, pp. 208–222, Jun. 2001.

[11] F. Resatsch, U. Sandner, J. M. Leimeister, and H. Krcmar, "Do Point of Sale RFID-Based Information Services Make a Difference? Analyzing Consumer Perceptions for Designing Smart Product Information Services in Retail Business," *Electron. Market.*, vol. 18, no. 3, pp. 216–231, Aug. 2008.

[12] R. T. Rust and P.K. Kannan, "E-service: a new paradigm for business in the electronic environment". Commun. ACM, 46, 6 (2003), 36–42.

[13] A.J. Sellen and , R. H. R Harper, The Myth of the Paperless Office. The MIT Press, 2001.

[14] F. Stalder, Politics of networked visibility, Acoustic Space, 10, (2011), 13–19.

[15] Universal design in Norway http://www.universal-design.environment.no/news/1063-norways-future-is-universal. Last accessed January 7, 2013.

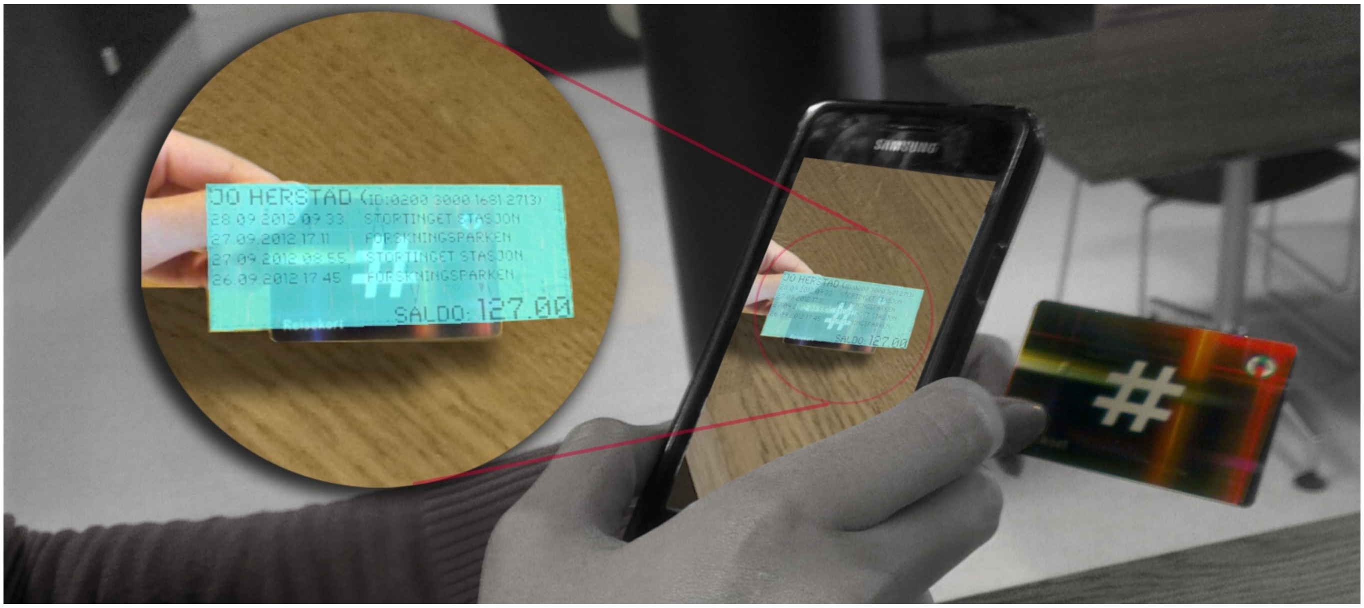


Figure 6. A 3D rendering of smart card ticket information using a NFC-enabled smart phone with ARTick.

TV Applications for the Elderly: Assessing the Acceptance of Adaptation and Multimodality

José Coelho, Tiago Guerreiro and Carlos Duarte
LaSIGE, Department of Informatics, University of Lisbon
C6 Piso3, 1749-016, Lisboa
Lisboa, Portugal
{jcoelho, tvjg, cad} @di.fc.ul.pt

Pradipta Biswas, Gokçen Aslan and Pat Langdon
Computer Laboratory, Department of Informatics,
University of Cambridge
15 JJ Thomson Avenue, Cambridge CB3 0FD
Cambridge, United Kingdom
{pb400, ga283, pml24} @cam.ac.uk

***Abstract*— Current TV applications present uncountable challenges to an elderly user and these are prone to increase, working as a vehicle of exclusion. The GUIDE project aims at improving the elderly experience with present TV applications by deploying interfaces adapted to their abilities and preferences. We do so by building the interface based on a user model and providing new ways to interact with the TV (multimodality), and tailoring the UI to the users' abilities, needs and preferences. In this paper, we assess concepts of the GUIDE framework with particular focus to the User Initialization Application (UIA), an interactive application able to build the aforementioned user model. We report an evaluation with 40 older users from two countries (UK and Spain). Results show that the UIA is able to create adequate profiles and that the users are able to positively observe the adaptations. Further, novel ways of interacting with the TV were also successfully evaluated as the users tended to experiment and rate positively most alternatives, particularly Speech and Tablet interaction.**

Keywords-accessible applications, elderly, multimodal, simulation, GUIDE

I. INTRODUCTION

The rapid increase of new TV-based systems and applications excludes users with certain impairments or different levels of technology awareness and expertise from accessing the same information as others do. By 2050 elderly people will represent 26% of the developed countries' population. Living in this constantly evolving world is getting harder, leaving them no time to adapt to modern technology from which, sooner or later, their well-being and social inclusion will dramatically depend. Having this type of technology adapt to each user without requiring too much learning and giving them similar opportunities as everybody else is the only way to prevent this. However, solutions being studied focus more on the use of assistive devices to help or guide the interaction than on the design of inclusive systems. We can offer valuable contributions for making this type of systems and applications more accessible by focusing on modern research in multimodal interactive systems, leading to less effort from elderly users, and offering them more natural ways of interaction.

This paper focuses on how the GUIDE project concentrates on the integration of several natural modalities and devices towards the inclusive use of TV-based applications in the user's home. Special attention will be given to the understanding of how elderly people can interact naturally with this type of technology. With this in mind, and after presenting related work, we will show how the GUIDE framework is built around keeping relevant user information stored and accessible in a User Model component. This allows adapting UIs, based on the specific user characteristics and preferences as well as on a context model. We will then focus on how a novel approach called User Initialization Application (UIA) was implemented. This process allows discovering each user's relevant characteristics for the system to adapt in the most appropriate and understandable way. At the same time, the UIA instructs the user on how to make use of the most suitable interaction possibilities available. We complement and attest these approaches with the description and results of a user study consisting of a technical trial with both the UIA and a TV based realistic Electronic Program Guide (EPG) application. We present results of the evaluation of the UIA capability to discover user characteristics and assign user profiles. With the EPG application we assessed users' modality preferences and evaluated the adaptation capabilities of the framework. The paper main contributions are the results of user studies conducted in two different countries (Spain and UK) with 40 participants, and the evaluation of novel approaches on adaptability and multimodal interaction of TV-based systems supported by the UIA application and GUIDE's multimodal and adaptive capabilities.

II. RELATED WORK

Elderly users are getting detached from technology and consequently from the modern world characterized by innovation in interaction and communication standards. In a society where people live longer, it's common to see elderly people living alone and struggling to adapt to all the technological advances [13]. Although older people are not generally considered to have disabilities, the natural ageing process carries some degenerative ability changes, which can include diminished vision, varying degrees of hearing loss, psychomotor impairments, as well as reduced attention, memory and learning abilities [20]. However, problems learning to use and engage with interactive technology are not confined to physical and cognitive factors [11]. In recent investigation, it is clear that even if computer based systems are a positive influence on the lives of older people [12], they tend to reject standard technologies [8] or technologies too

 ISBN: 978-1-61208-250-9

difficult to use, that look bad, make them appear older, or that do not appear to have net gain effect for them [7].

For these reasons, Information and Communication Technology (ICT) must be adjusted to ensure that elderly users are not disadvantaged when using it [7, 8]. This accommodation can only be done by realizing the consequences associated with ageing and offering natural and efficient ways of interacting with both new and well-known technical systems.

Multimodal interfaces aim to provide a more natural and transparent interaction to users. They have been able to enhance human-computer interaction (HCI) in many ways, including: User satisfaction: studies revealed that people favor multiple-action modalities for virtual object manipulation tasks [18], and that about 95% of users prefer multimodal interaction over unimodal interaction [18]; Robustness and Accuracy: “using a number of modes can increase the vocabulary of symbols available to the user” leading to an increased accessibility [16]. Oviatt stated also that multiple inputs have a great potential to improve information and systems accessibility, because by complementing each other, they can yield a “highly synergistic blend in which the strengths of each mode are capitalized upon and used to overcome weaknesses in the other” [19]; Efficiency and Reliability: Multimodal interfaces are more efficient than unimodal interfaces, because they can in fact speed up tasks completion by 10% and improve error handling and reliability [17]; and Adaptivity: Multimodal interfaces offer an increase in flexibility and adaptivity in interaction because of the ability to switch among different modes of input, to whichever is more convenient or accessible to a user [16]. For all these reasons, multimodal interfaces have the potential to increase ICT applications’ accessibility for elderly users. Introducing elderly users to multimodal solutions makes possible a familiar and natural step through the accommodation necessary to ensure that older users are not disadvantaged when using new technologies like the ones based on TV and STBs [13, 21].

Interface personalization is mainly explored in the domain of content personalization and developing intelligent information filtering or recommendation systems based on user profiles. In most of those systems, content is represented in a graph like structure and filtering or recommendation is generated by storing and analyzing users’ interaction patterns. Little research has been done beyond content personalization. A few significant projects on interface personalization are the SUPPLE project[10], the Lumiere Project [14] and the AVANTI project [22] for people with disabilities. The SUPPLE project personalizes interfaces by changing layout and font size for people with visual and motor impairment and ubiquitous devices. However, the user models do not consider visual and motor impairment in detail and thus work for only loss of visual acuity and a few types of motor impairment. The Lumiere project uses an influence diagram in modeling users. This records the relationships among users’ needs, goals, background etc. The AVANTI project provides a multimedia Web browser for people with light or severe motor disabilities, and for blind people. It distinguishes personalization into two classes: static adaptation which is personalization based on user profile and dynamic adaptation that is personalization following the interaction pattern with the system. However, the Lumiere project does not generalize their personalization mechanisms for other applications and the AVANTI project only addresses a small segment of disabilities for a particular application. The lack of a generalized framework for personalization of users with a wide range of abilities affects the scalability of products.

III. The GUIDE Project

As a multimodal adaptive system, GUIDE offers several modalities of interaction coordinated by the system core, through fusion and fission modules that function based on a User Model [4], which represents the user, and a Context Model, which represents the context (environment, devices, etc.) the user is in. Input modalities are based in natural ways of communication for humans: speech and pointing (and gestures). Complementary to these modalities and being based in a TV environment, the system also supports the usage of remote controls (both traditional and endowed with gyroscopic capabilities) and other devices capable of providing haptic input or feedback. GUIDE incorporates four main types of UI components: visual sensing and gesture interpretation; audio; remote control; haptic interfaces and a multi-touch tablet. In what concerns the output modalities, the framework integrates the following: video rendering equipment (TV); audio rendering equipment (speakers); tablet supporting a subset of video and audio rendering; and remote control supporting a subset of audio rendering and vibration feedback. The tablet, used as a secondary display, may be used to clone the TV screen or complement information displayed on the TV. In addition to the application’s interface, the framework is capable of rendering a 3D avatar. This is expected to play a major role for elderly acceptance and adoption of the GUIDE system, being able to perform non-verbal expressions like facial expressions and gestures and giving the system a more human like communication ability. Both input and output modalities can be used in a combined manner to enrich interaction and reach all user types. In order for the UI to be adapted to the user’s preferences and abilities, the interface elements are highly configurable and scalable (vector-based). Graphical properties like size, font, location and color are some attributes needed to ensure adaptability. Other modalities’ properties, like sound volume, are also configurable. The GUIDE User Model, responsible for the dynamic adaptation of the system, will be described in a later section.

A previous study was conducted last year [5] with the goal of identifying the most relevant UI component configurations and user impairments for elderly users. This was particularly important for deriving a list of relevant topics and metrics to be considered in the development of accessible applications for these users. Closely connected to the new study are the conclusions related with the way elderly prefer and perform multimodal interaction in TV-based systems. However, in the previous study the application used was not a realistic TV-application. Thus, a more contextualized evaluation can be done. Additionally, the process of user characterization and clustering is now focused.

IV. THE USER MODEL

The GUIDE user model maps users' functional parameters to interface parameters. It was developed using a simulator [4] and calibrated through the user study [5] discussed in the previous section. The simulator consists of detailed models of visual and auditory perception, cognition and motor action. The simulator can show the effects of a particular disease on visual functions and hand strength metrics and in turn their effect on interaction. Using information from more than 100 users from three different countries (Spain, UK and Germany), collected with an extensive survey focusing a wide array of characteristics, we have selected a set of variables that are relevant to the user model and statistically significantly different among clusters ($p<0.01$). We separately clustered these data for visual, cognitive and motor abilities of users using k-means clustering. Table 1 shows the cluster centers.

Table 1: Cluster Centers

Visual[1]	A	B	C
CS1 (Contrast Sensitivity on Pelli Robson Chart n. 1)	0.15	1.75	1.46
CS2 (Contrast Sensitivity on Pelli Robson Chart n. 2)	0.15	1.84	1.56
CS3 (Contrast Sensitivity on Pelli Robson Chart n. 3)	0	1.79	1.37
Cognitive	A	B	
TMTSEC (Time to complete Trial Making Test in sec)	45.97	115.3	
DIGIT_SY (Result on Digit Symbol Test)	45.93	23.71	
Motor	A	B	C
GS (Grip Strength in Kg)	16.22	25.02	58.68
ROMW (Active Range of Motion of Wrist)	71.28	51.58	65.67

Following this, we ran the GUIDE simulator [4], taking the parameters of each cluster centre for configuration, and generated recommendations for each cluster. Individual users were assigned to the recommendations based on their cluster memberships. For the present set of users we identified the following three profiles: Profile A: No adaptation required; Profile B: Increase button spacing (Mobility Impaired); Profile C: Increase button spacing + change Color Contrast (Mobility Impaired + Color Blind).

The GUIDE User Model predicts three sets of parameters: UI parameters for the Multimodal Fission Module, Adaptation Code for the Input Adaptation Module and Modality Preference for the Multimodal Fusion Module. The rules relating the users' range of abilities with interface parameters were developed by running the simulator [3, 4] in Monte Carlo simulation.

We began by selecting a set of variables to define a Web based interface. These parameters include: Button spacing: Minimum distance between two buttons to avoid missed selection; Button Color: foreground and background color of a button; Button Size: The size of a button; Text Size: Font size for any text rendered in the interface; Cursor Type: The shape and color of the cursor.

The user model predicts the minimum button spacing required, from the users' motor capabilities and screen size. The simulation predicts that users having less than 10 kg of grip strength or 80º of Active Range of motion of wrist or significant tremor in hand produce a lot of random movement while they try to stop pointer movement to select. The area of this random movement is also calculated from the simulator. Based on this result, we calculated the radius of the region of the random movement and the minimum button spacing is predicted in such a way so that this random movement does not produce a wrong target selection. Regarding the other parameters, the UIA takes user preferences for color, text size and cursor type. The user model stores these preferences. However if a user has color blindness it recommends an appropriated foreground and background color. The adaptation code aims to help users while they use a pointer to interact with the screen through the visual human sensing capabilities or the gyroscopic remote. So if the user has any motor impairment, the adaptation will remove jitters in movement through exponential average and then attract the pointer towards a target when it is near by using the gravity well mechanism [3]. Otherwise, the adaptation will only work to remove minor jitters in movement. The modality prediction system predicts the best modality of interaction for users. Though users are free to use any modality irrespective of the prediction, the fusion module uses this prediction to disambiguate input streams when there is more than one.

V. USER INITIALIZATION APPLICATION

The UIA is an introductory application that runs the first time a user initializes the system. When a new user is recognized, the UIA presents a step-by-step introduction of the system, acting as a tutorial on how to use the system and how to interact using the different modalities available. Another purpose of the UIA, as important as the one described before, is to expedite the user profiling procedure. It would not be feasible to ask each user of a mass market product to complete an extensive survey before using it. The UIA presents a much reduced set of questions and tasks to the user in order to allow the User Model to assign the user to one of the previously created profiles. Even though this does not allow for a profile perfectly fitted to the user, it is a good starting point for adaptation purposes, because the profiles described in the previous section were created from a large pool of representative users. Additional information collected during system usage is used to refine the user profile (run-time adaptation).

The tasks and metrics chosen for the UIA are the ones for which the resulting data is the most capable to assign the more appropriate profile to the user profile. They were selected from an analysis of the extensive survey data, taking into account the feasibility of gathering the data. For those instances where it was not feasible to gather the data in a living room environment, alternative sources were selected

[1] Each of these clusters is further divided based on presence of color blindness

and combined to estimate the required data. A description of these variables is listed below: Color Blindness: Plates 16 and 17 of Ishihara Test [6] as it may classify among Protanopia, Deuteranopia and any other type of color blindness; Dexterity: We estimated Grip Strength and Active Range of Motion of wrist from age, sex and height of users following earlier Ergonomics research [2]; Tremor: We conducted earlier a test involving a Tablet device in horizontal position, and estimated tremor from the average number of times users need to touch the screen to select small buttons. Details of the study can be found in a separate paper [4]. Additionally, other tasks were chosen with the purpose of allowing users to personalize the system, while being a hands-on tutorial regarding new modality interaction and feedback configuration. The most relevant ones are the following: Modality Introduction: Self-explanatory videos of how to interact with each modality, followed by "do-it-yourself" tasks; Button and Menu Configuration: Button size, and font and background color configuration; Cursor Configuration: Cursor size, shape and color configuration; Audio Perception: Hearing capabilities and preferences.

The UIA has a simple user interface, with a different screen for every task and metric identified above. Few buttons are presented per screen (preventing user confusion). Every screen preserves the same navigation model - an area with "next", "previous" and "repeat" buttons, and another visually distinct area for presenting information and requests. For every metric to be measured, tests are presented as simple questions about preferences. Also, for every modality available in the system, a video introducing its use is presented, followed by the possibility for the user to try it out. A virtual character accompanies the user through this process, offering explanations and assisting the user in the personalization. As the user goes through each task and preference setting, the UIA adapts itself to the preferences already manifested. For example, if user manifests preference for big, blue buttons with yellow text, all buttons will be presented with those settings from that moment onwards. It is worth pointing out that the results of our previous study are reflected in the UIA's design: high contrast colors, big, centered and well-spaced buttons, etc.

VI. STUDY DESCRIPTION

The study's main goal is to assess the acceptance of adaptation and multimodality by elderly users when interacting with TV based applications. We do this by addressing several questions related with the UIA, multimodal interaction and adaptation, and GUIDE in general. Regarding the UIA: first we want to measure the efficacy of this application in discovering the relevant characteristics of users and assigning user profiles; secondly, we want to evaluate how understandable the UIA is in terms of its goals and the instructions it provides; and finally, how easy it is for elderly to interact with this application, or if they would do it if it was part of their daily lives. Regarding multimodal interaction, we want to assess which modalities are the most used by elderly in a realistic TV interaction scenario. In what relates to adaptation, the goal of this study is to understand if users perceive it and if they are more satisfied using UIs adapted to their characteristics than non-adapted versions of the same applications. Lastly, regarding GUIDE in general, we want to measure its acceptability and if elderly users perceive improvements in their quality of life just by using this system.

A. Development of a Realistic EPG

Similarly to the UIA, we used the knowledge gained in the first study to develop an Electronic Program Guide (EPG) application and focused on the engagement between users and a realistic TV interaction scenario. By developing this application we wanted to validate previously developed notions about elderly users interaction. For instance, do users favor alternative and multimodal ways of interaction with the TV or when they are in the presence of a realistic application they prefer the traditional interaction devices, like a remote control?

The development of this application started with mockup designs, which were then reviewed and approved by experts in the development of TV and STB applications. Later all elements were implemented using HTML, JQuery, CSS and JS languages. The EPG had real information about channels and respective schedules and shows, to confer more realism to the application. Adapted versions were developed to fit each user profile. Output modalities, including the Virtual character, were made available and selected according to the user profile.

B. Participants (Pre-Survey)

We recruited 40 elderly people (24 female and 16 male) with different age-related disabilities. Users were recruited in two countries, with 21 participants (14 female and 7 male) being recruited in Spain and 19 participants (10 female and 9 male) in the UK. The average age was 70.9 years old and the different user profiles were assigned to the participants in the following manner: 14 users with profile A, 22 users with profile B, and 4 users with profile C. As the non-adapted version of the EPG was developed already addressing accessibility issues, each of the profiles reflected few adaptations. All users participated voluntarily and all activities involved in this study were safeguarded from the ethical point of view.

C. Apparatus

The study was conducted in two locations (Spain and UK). Efforts were directed to create similar environment and technical conditions in both labs. Trials were conducted by usability experts. Users were given freedom to interact (the trial conductor would only intervene when really needed, or user asked for help). In what concerns the technical setup and specification, different modalities of interaction were configured: pointing resorted to the use of a Microsoft Kinect; for speech recognition we used the Loquendo SR engine; a simplified remote control, with less buttons than traditional ones and capable of controlling pointer coordinates using a gyroscopic sensor was made available; an iPad was used for tablet interaction; and a full 1080p HDMI TV with integrated speakers and a 32'' screen was used for visual and audio output. User interactions and answers were video recorded.

D. Design and Analysis

We used a within-subjects design where all users ran the UIA and were evaluated in both adapted and non-adapted settings. The order in which they performed both versions of the EPG was randomized to counteract learning effects. Qualitative analysis was retrieved from the pre, intermediate and post-questionnaires. Quantitative data was retrieved from the UIA (user profile and interface preferences). Herein, we discarded quantitative measures like trial errors and time as the trials followed a semi-supervised methodology: the participants were motivated to perform the tasks on their own but they were free to ask questions when they felt lost. Wilcoxon Signed rank tests were used in comparisons to subjective measures between both versions of the EPG. Whereas to binomial measures, Mcnemar's was performed. Cohen's Kappa was used to assess the inter-reliability of the profile ratings.

VII. RESULTS

The goal of this study was to make a preliminary assessment of our adaptation approach for interacting with the TV. We focused our attention on how well we are able to interactively retrieve information about the user and generate adaptations accordingly. We start by analyzing how well our UIA is able to profile the users along with an assessment of the initialization interface itself. Then, we analyze how the users interacted with the different modalities and their acceptance of the multimodal concept. To end, we evaluate subjective acceptance towards adaptation and GUIDE in the overall.

A. Discovering Elderly Profiles with UIA

Our take for adaptation relies on a User Model fed by the UIA. All participants in our study performed both the pre-survey and the UIA. Twenty-nine out of forty profile assessments were performed similarly by the two methods (74%). The interrater reliability between the profiles assigned with the pre-survey and the UIA was found to be Kappa = 0.58 ($p < .0.001$), revealing a moderate agreement [15]. It is relevant to notice that the UIA enables the user to input preference values, something that goes beyond ability profile. This is likely to explain part of the mismatch (e.g., a user with no visual impairments is likely to prefer a higher contrast button when he is confronted with such an hypothesis). Another source of uncertainty may be the under-statements by part of the users in the pre-survey. Indeed, in a questionnaire it is likely that part of the users fail to acknowledge some limitations while they clearly state them when confronted with an interface with options to surpass it. A deeper understanding of the mismatches that are not created by these observed flaws can only be retrieved in a more extensive evaluation by analyzing how both methodologies enable the users to improve performance.

B. UIA evaluation by the Elderly

As mentioned before, we have reproduced a realistic EPG and have focused our attention on improving its overall accessibility. Conversely, we have included the UIA, a component the users are not used to. It is important to assess how the users see this component and if they are willing to use such a thing to improve their performance.

The participants took between 12 and 37 minutes to complete the UIA (M=22.8, SD=5.9). Once again, although they were discouraged to engage in long dialogues the participants were free to express their opinions and doubts during the UIA which increased the time to finalize the process. The UIA classified 16 people as profile A, 20 as profile B, and 4 as profile C. Table 2 presents the subjective ratings given by all the participants to the questions posed. Regarding the understanding of the purpose of the UIA (Question 1), 9 out of 40 (22%) did not understand the purpose of the UIA. This indicates that such a process should be better motivated or else it will be likely ignored by the users. In line with this, 11 out of 40 (28%) stated they would skip the process if they had the system at home (Q2). Five participants stated to find the process too long while four other were neutral about it (Q3) All the remaining thought it was neither too long nor tiring. Most users (35) thought the UIA was easy to follow and understand (Q4). Regarding the adaptations felt during the UIA (Q5), 26 participants stated to have noticed them. This is easily explained as 16 participants were classified as profile A which means they had little or no adaptations done during the UIA. In sum, the users seem positive towards the UIA (Table 2) although it is clear that it should be well motivated and accompanied.

Table 2: Subjective ratings of the UIA

Question about the UIA	Median	IQR
Have you understood why we do the UIA? [1 - Yes ; 2 - No]	1	0
If you have had the system at home, would you go through it or skip it? [1 - Would do it ; 2 - Would skip it]	1	1
Do you think the UIA is too long? [1- Yes;2 - Neutral;3 - No]	3	0
Were the instructions easy enough to understand? [1 - Yes; 2 - No]	1	0
Did you notice any changes in the application while you were using it? [1 - Yes ; 2 - No]	1	1

C. Evaluating Multimodal Interaction

One of GUIDE's main concepts relies on offering a set of modalities to the elderly population to fit their different profiles and abilities. In this study, in the Adapted setting we allowed the participants to perform the tasks with any of the previously described modalities. After each task, the users were asked about their preferred modality. Figure 1 presents their preference count after each task. Overall, it shows that both Speech as Tablet interaction were seen as positive improvements for interacting with the EPG. Conversely, the standard remote is seen by a part of the users as a safe option which they are not willing to trade easily. The gyroscopic remote was mildly used while pointing at the screen was the least preferred unimodal option. The users were also able to use Pointing and Speech together but this option was not revealed as interesting. This may be due to the complexity of using more than one modality and the reduced timespan of training. Further, the simplicity of the tasks is likely to be fitted with simpler selection approaches while a multimodal option is directed at more complex tasks. Nonetheless, given the population and the overall goal of simplifying processes and the aforementioned trends, it seems reasonable to suggest that a variety of unimodal interfaces seems to be adequate to fit the users' needs and preferences. Interestingly enough, when explained the modalities available in the adapted version, most users stated that they would stick with standard remote. Conversely, during the trials they showed interest in using other modalities.

Looking at the order of the trials, those that ran the adapted version first seem to be eager to try out different modalities, particularly the gyroscopic remote, tablet and speech, while those that performed the tasks previously with the Non-Adapted version, use a more conservative approach in the Adapted setting, giving preference to Speech followed by the standard remote and the tablet and ignoring the remaining.

In both scenarios, about half the users shifted their preference during the Adapted trial while the other half sticked with their initial preference (first task). Preferences also diverge between users from different profiles. Participants from Profile A (no adaptation) showed a predilection for both the standard remote and speech and a mild preference for using the tablet. All the other modalities were ignored. As to Profile B (medium adaptation level), participants showed a less consistent selection. The tablet was the most preferred modality but all the other unimodal approaches were seen as useful by part of the users. As to Profile C (high adaptation level), consistency was once again revealed with preferences going for both remotes, pointing and speech

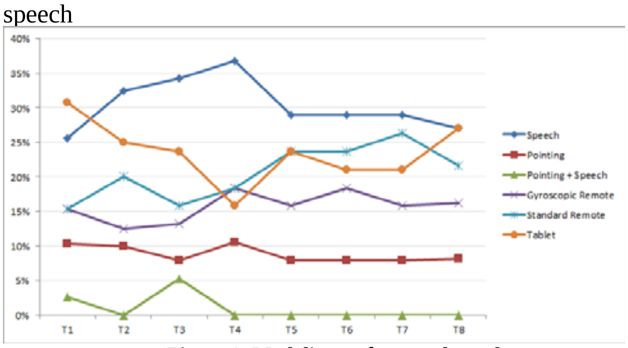


Figure 1: Modality preferences by task

This first analysis on how older people resort to non-conventional modalities reveals acceptance of the underlying multimodal concepts: results suggested that the population does not reject the usage of other control interfaces if these are adapted to their profiles and the task in hand. Further, Age did not show a significant correlation with acceptance nor any observable tendency on the set of preferred modalities, suggesting that this overall acceptance is pervasive to the age groups in our sample.

D. Evaluating Adaptation

Subjective acceptance to EPG adapted and non-adapted versions was evaluated through a statement to classify using a five-point Likert scale (1-the system has not supported me at all; 2-the system has supported me only in some parts of the task; 3-undecided; 4- the system has supported me in almost all the tasks; 5-the system has supported me at every moment). A Wilcoxon Signed Rank Test revealed a statistically significant minor difference between the EPG adapted (Mode=4, IQR = 1) and non-adapted versions (Mode=3, IQR=2), $z=-1.665$, $p<.1$, with a medium effect size ($r=0.37$). Although not significant statistically, subjective acceptance of adaptation showed to be slightly higher by participants in profile C, suggesting that the adaptations (more noticeable) improved the relation between the participants and the interface. In-detail acceptance comparisons of icon and text properties (evaluated as a dichotomous response to the suitability of the property) in both adapted and non-adapted settings revealed no significant differences. Apart from the large number of users with no adaptation, the baseline EPG was already designed as an accessible version. This may be the reason for the participants to rate the non-adapted EPG as adapted to their needs as well.

E. Evaluating GUIDE concepts

The adaptation and multimodality concepts pervade the GUIDE project. These preliminary trials seeked to assess how older people would react to such an adaptable and flexible system. While the aforementioned behaviors suggested that both adaptation and a wide coverage of control interfaces were positively seen and felt by the participants, we questioned them directly about their opinions on GUIDE. To this end, we performed a self-rating post-questionnaire to assess their overall opinion about the system (5-point Likert scale) and their Behavioral Intention to Use the System (7-point Likert scale).

Table 3: Subjective ratings to GUIDE, Median[IQR}

Overall opinion about the system	Rating [1-5]
Overall, I am satisfied with how easy it is to use	4 [2]
I am able to efficiently complete the tasks using the system	4 [2]
I feel comfortable using the system	5 [1]
It was easy to learn to use it	4 [1]
Whenever I make a mistake, I recover easily and quickly	4 [2]
it is easy to find the information I needed	4 [1]
The interface of the system is pleasant	5 [1]
Overall, I am satisfied with the system	5 [1]
Behavioural Intention to Use the System (BI)	**Rating [1-7]**
Intend to use GUIDE in the next semesters if I have access to it.	6 [2.5]

Overall, all items were highly positively rated by most users (table 3). Regarding the opinion about the system, the overall satisfaction with it, the comfort and pleasure it guarantees were consistently highly rated as shown by the median and low dispersion. The ratings with slightly larger dispersions were about the easiness, efficiency and

 ISBN: 978-1-61208-250-9

recovering of errors of the system, although still on the positive end of the rating scale (Median 5, IQR = 2). This can be explained by technical errors with a small part of the users regarding both the speech recognizer and the tablet device (temporarily unavailable wireless access). These users were unable to complete the tasks as they desired which translated in the mentioned lower ratings. This, although it can be explained by a technical glitch, is a relevant lesson learned as an unexpected, even localized, flaw, particularly in an adoption phase, is prone to damage the user's relationship with the system as a whole.

When asked about their Behavioral intention, users were also very positive which can be observed by an overall median value of 6 in 7 regarding intention of use. Here, the dispersion is a little larger. Five participants, although enjoying and understanding the benefits of GUIDE, argued to be happy with their systems at home, and therefore were neutral about a change. Only two participants showed to be unwilling to adopt GUIDE: one of the users showed preference to use the tablet for all tasks and experienced the aforementioned technical glitches which can explain his position; the other user did not have a TV at home, did not show an interest in having one, and presented an overall attitude towards technology ("I do not like technology"). Interestingly enough, he did not resort to conventional interaction modalities.

VIII. Discussion

Upon analyzing the UIA process and its impact on adaptation along with the usage of the GUIDE system and its underlying concepts, we answer our research topics as follows:

A. Deriving a suitable user adaptation profile through the UIA.

The UIA aims at creating a user profile by performing a simple set of questions and interactive tests. Results showed that the UIA is able to match profiles obtained with an extensive survey in 74% of the cases. Further, the UIA showed to be more realistic than its paper-based counterpart as data is likely to be more accurate when the users are faced with their limitations rather than just being questioned about them. Moreover, the UIA gives space for preference and subjectiveness. In sum, we consider that adapted TV applications based on simple initialization profiling are feasible and likely to improve over traditional methodologies.

B. Acceptance of the UIA.

The UIA took over 12 minutes, averaging around 23 minutes. This amount of time can be discouraging for an elderly user if the benefits are not clear. Taking in consideration that it is supposed to be ran only once, the participants showed to be very positive about it. This is supported by the almost general understanding of the purpose of the UIA: they understood the benefits of such an application and perceived the adaptations during the process. Most participants (35) considered the application easy to follow which indicates that although the concepts underlying the creation of the user model are complex, the interface to generate it is not.

C. On perceived adaptation.

Our analysis on adaptation is restricted to the subjective understanding and acceptance of the created profiles and consequent adaptation. Results showed that participants perceived the adaptation both during the UIA and the adapted EPG tasks. Also, those that were subject to adaptations rated the adaptive version as an improvement over the non-adapted one. The baseline EPG was already an improved accessibility-wise version over traditional EPGs, a fact that may have reduced the impact of the adaptations in such a short term evaluation. The impact of these adaptations needs further longitudinal evaluations supported by quantitative measurements to be further assessed. Nonetheless, the participants showed to be positive about the adaptations, which is relevant as a requirement for adoption, particularly in an elderly population.

D. On Multimodality.

It is commonplace to underlook the elderly population as one that is attached to traditional methods and unwilling to adopt new technologies. We acknowledge that the adoption of new technologies by the elderly is not straightforward and needs to be supported and accompanied. In this study, the participants showed to be eager to try new methodologies and experience their benefits over traditional counterparts. Speech recognition and interacting with the Tablet were tried during the tasks by most users and together achieve over 50% of the participants' preference. On the other hand, more conservative users, although resorting to other modalities when they saw fit, selected the standard remote as their main control interface. Providing an enriched set of modalities showed to fit the different user profiles. In that sense, multimodality seems to provide the flexibility argued in our motivation. Conversely, the unique multimodal option provided (Pointing + Speech) failed to prove the combination of modalities as useful for the target population. This may be due to the difficulty (easy) of the tasks in hand.

E. Overall acceptance of GUIDE.

Participants showed an overall positive acceptance towards GUIDE. Some participants stated that their EPGs were difficult and would desire a simpler system. They saw it in GUIDE. The avatar in the UIA was seen as a friendly helper and the overall usage of the EPG was considered simple and comfortable. Moderate opinions were given only in extreme cases of technology rejection (1 user) and technical glitches that disabled the users from completing the tasks as desired (inability to use their preferred interfaces). Participants stated to be interested in using GUIDE if it would be made available

IX. Conclusions

New interaction paradigms, supported by new modalities and applications, are transforming a classical appliance that is the TV. If not handled properly, this transformation can increase the access barriers to TV content for elderly users.

In this paper, we assessed several of the proposals that the GUIDE project puts forward in order to increase the accessibility of TV applications. GUIDE aims to provide application developers with a multimodal adaptive framework and a set of functionalities that will increase their products' accessibility, without demanding major changes in their development process. The assessment was based on a user trial, with 40 participants from two different countries.

Being an adaptive framework, it relies on knowledge and information about users. We described how the GUIDE User Model was built and how it integrates with the framework. Essential for the integration, is the UIA, a process that streamlines user profile identification, based on short number of tasks and questions. We present an assessment of the efficacy of this process, concluding that it is possible to reliably identify user profiles, while also recognizing ways in which to further improve the process. From the user's point of view, the process motivation was understood, and it was considered easy enough, although also here we were able to find ways to improve it.

Being a multimodal framework, it makes use of several input and output modalities. The paper assessed the usage of these modalities in the context of a realistic EPG application. Conversely to what could be expected, the traditional remote control was not the most favored modality, being surpassed by speech and tablet as preferred modalities. Effects of adaptation on this application, in spite of an already accessible non-adapted version and the limited amount of interaction time with it, achieved a statistically significant minor difference between subjective acceptance of adaptive and non-adaptive versions.

These results, together with the positive acceptance of the GUIDE concepts and their expected impact in the quality of life of its users, validate the approach followed so far and pave the road for the project's future developments, which will be verified in a longitudinal trial for better assessing the effects of adaptation

A. *Relation with the Previous Study.*

Regarding the use of modalities, speech interaction was singled out as the most attractive modality. In the first study, a Wizard-of-Oz approach was used to replace the speech recognition engine, and we questioned how that might have contributed to the results. In this follow-up study, where a speech recognition engine was used, we can see that speech remains the preferred modality, overcoming any technical issues that might be raised. It seems safe to say that speech plays an important role in promoting the adoption of these systems, and efforts to ensure its adequate operation are justified by the satisfaction it provides users with. Tablets, although not fully integrated with the system in the initial study, collected a positive response from participants, with 92% of them considering interacting with a TV using the Tablet. This tendency was confirmed in this study, with Tablet interaction being the second most used modality to interact with the EPG. Finally, regarding the clustering process, by increasing the number of users available we have updated the profiling process, which resulted in a more accurate representation of the users' characteristics and a more precise identification of the relative importance of each variable. For example, contrast sensitivity is now more relevant than the capability of seeing at distance or at night, and grip strength is now more relevant than any other motor related variable.

REFERENCES

[1] Anderson J. R. and Lebiere C. "The Atomic Components of Thought", Lawrence Erlbaum Associates, 1998.

[2] Angst F. et. al.., Prediction of grip and key pinch strength in 978 healthy subjects, BMC Musculoskeletal Disorders 2010.

[3] Biswas P. and Langdon P. (2012) Developing multimodal adaptation algorithm for mobility impaired users by evaluating their hand strength, IJHCI, Taylor & Francis.

[4] Biswas P., Langdon P. & Robinson P. (2012) Designing inclusive interfaces through user modelling and simulation, IJHCI, Taylor & Francis, Vol; 28, Issue 1, 2012

[5] Coelho J., Duarte C., Biswas P. and Langdon P., Developing Accessible TV Applications, Proceedings of ASSETS2011

[6] Colour Blindness Tests 2008. Available at: http://www.kcl.ac.uk/teares/gktvc/vc/lt/colourblindness/cblind.htm, Accessed in 12th August February, 2008

[7] Czaja, S. and Lee, C. (2007). The impact of aging on access to technology. Univ.Access Inf. Soc. 5 (4), 341-349.

[8] Dickinson, A. and Gregor, P., 2006. Computer use has no demonstrated impact on the well-being of older adults. Intl. Journal of Human Computer Studies. 68, 744-753.

[9] Duffy V. G. "Handbook of Digital Human Modeling: Research for Applied Ergonomics and Human Factors Engineering." Boca Raton, FL, USA: CRC Press, 2008

[10] Gajos K., et al.Automatically generating user interfaces adapted to users' motor and vision capabilities. UIST 2007.

[11] Hanson, V. (2009) Age and web access: the next generation. Proceedings of W4A 2009

[12] Harley D. & al. (2009). Age Matters: Bridging the Generation Gap through Technology-Mediated Interaction. In Proc. of CHI.

[13] Holzinger, A., Ziefle, M. and Rocker, C. Human-computer interaction and usability engineering for elderly (HCI4AGING). In Proceedings of ICCHP'10.

[14] Horovitz, E. and colleagues (2010), Microsoft Research, The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users.

[15] Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics 33:159-174.

[16] Oakley, I., Brewster, S. A., and Gray, P. D.: Solving multi-target haptic problems in menu interaction. In Proc. CHI'01.

[17] Oviatt, S. Multimodal interactive maps: Designing for human performance. Human-Computer Interaction, 1997. 93-129

[18] Oviatt, S. L., DeAngeli, A., and Kuhn, K. Integration and synchronization of input modes during multimodal human-computer interaction. CHI '97.

[19] Oviatt, S.L.: Mutual Disambiguation of Recognition Errors in a Multimodal Architecture. CHI'99.

[20] S. Kieffer, A. et al., "Towards StandardizedPen-Based Annotation of Breast Cancer Findings", Proceedings HCII09.

[21] Sharon Oviatt, Trevor Darrell, and Myron Flickner. 2004. Multimodal interfaces that flex, adapt, and persist. Commun. ACM 47, 30-33.

[22] Stephanidis, C. et al: Adaptable and Adaptive User Interfaces for Disabled Users in the AVANTI Project. S.Triglia et al. (Eds.): IS&N'98, 153-166, 1998

Identifying Cross-Platform and Cross-Modality Interaction Problems in e-Learning Environments

André Constantino da Silva
Institute of Computing (PG)
UNICAMP, IFSP
Campinas,Brazil, Hortolândia,Brazil
acsilva@ic.unicamp.br

Fernanda Maria Pereira Freire
NIED
UNICAMP
Campinas, Brazil
ffreire@unicamp.br

Heloísa Vieira da Rocha
Institute of Computing, NIED
UNICAMP
Campinas, Brazil
heloisa@ic.unicamp.br

***Abstract*—Web applications and sites are designed to use keyboard and mouse as input devices and a medium resolution screen as output device. Mobile devices, such as smartphones and tablets, have enough computation power to render Web pages, allowing browsing the Internet. But, their main interaction style is touching style that was not usually considered in the Web applications design. Changing the platform or interaction style can lead to interaction problems. To study these problems, we investigated the use of TelEduc, an e-Learning environment designed to Internet and to be used with keyboard and mouse, in two touchscreen devices, a smartphone and a tablet. Some problems are usability problems and do not have relation with the platform or modality, but other problems are related to the platform or modality changing.**

Keywords-Mobile devices and services; Interfaces, interactions and systems for distance education; Interface evaluation; Usability testing and evaluation;

I. INTRODUCTION

Devices, such as smartphones and tablets, are becoming increasingly popular; most of them have touch screen displays, access to the Internet and enough computing power to process Web pages. So, Web sites and Web applications, initially developed to be used with keyboard, mouse and a medium size display, are been accessed by small touch screen devices.

One kind of Web applications is e-Learning environments, as Moodle [1], SAKAI [2] and TelEduc [3], which are applications with tools to support teaching and learning activities though the Web. These tools allow users to create content, communicate with other users and manage the virtual space.

These environments take advantages of the Web to offer content with text, images, audios and videos in a hypertext document. Tools like chat, forums, portfolios, repositories are widely used, and tools that explore the audio and video resource to user communication, such as instant messenger and video-conferences, are becoming common among the environments.

Since smartphones and tablets are easy to carry, have autonomy for hours and Internet access, the e-Learning environments´ development teams are building solutions to provide access on mobile devices. Three kind of solution are emerging: specific device application, Web site specific for mobile devices, and improve the Web site for mobile and desktop access.

Browsing Web site using another device (last two solutions), the user can deal with some problems related with the platform changing (e.g., from desktop´s browser to smartphone´s browser), but some problems happen due the interaction style changing (e.g., keyboard, mouse and medium screen to a small touchscreen). We call these problems as cross-platform interaction problem and cross-modality interaction problems, respectively.

So, which cross-platform and cross-modal interaction problems users deal? How to categorize a problem? Thinking about this question in the e-Learning context, we developed this work. Though a user interface analyze by a specialist using a smartphone, a tablet and a desktop, we found some problems to browsing TelEduc using two touchscreen devices: a smartphone and a tablet. These problems were classified into categories: cross-platform problem, cross-modality problem, and platform and modality- independent problem.

The next Section describes related works. Section III presents the TelEduc Project with a brief historical view, the tools and features of TelEduc e-Learning environment. Section IV shows the material and method adopted. Section V describes some identified interaction problems, and, in the Section VI, these problems are classified into the three described categories. Section VII presents conclusion and future works.

II. RELATED WORK

The e-Learning environments´ development teams are building solutions to provide access on mobile devices. Three kind of solution are emerging: specific device application; Web site specific for mobile devices; and improve the Web site for mobile and desktop access.

Building specific device application allows designing a suitable user interface for the device and taking advantages of smartphone´s features, such as touchscreen and camera, but needs develop an application for each mobile platform, so to be developed needs specific knowledge programming team and increases the code lines number to maintain. Moodle community offers the Moodle App [4] and Moodbile

[5], two native mobile applications with versions for the most popular smartphone´s platforms.

Moodle, since version 2.1, offers a Web site specific to mobile devices, an example for the second type of solutions for access e-Learning environments in mobile devices. Building a specific Web site to mobile device allows designing a suitable user interface for mobile devices taking account some common characteristics, such small touchscreen, but depends of the browser to access some platform features, such GPS, and increases the code lines number to maintain too.

The latter solution considers that smartphones and tablets have enough computational power to render Web pages and to do some adaptation if it is necessary, and offer the same user interface for any device. To design this kind of user interface it is necessary to do some usability studies to found barriers or user interaction problems. Disadvantages of this solution are to depend of browsers to use the mobile features and the difficult of consider many interaction styles in the same user interface.

This solution can start from a user interface design model for desktop and be improved to consider mobile devices. So in the initial design was designed thinking the user will interact by keyboard, mouse and medium size display and, allowing users accessed these applications on mobile devices, there is an increasing of interaction styles number, such touchscreen. With the interaction hardware changing the user deals with new interaction problems. Shrestha [6] points out some problems when the users try to use mobile devices to do specific tasks into Web sites designed to desktop, so the mobile Web browsing experience needs to be improved to a more mobile friendly Web site and some mobile browser improvement (here we consider browser as one platform characteristic).

Shrestha [6] considered mobile devices equipped with joystick and a small screen. Maurer *et al.* [7] did some usability studies using touchscreen mobile devices and desktop for browsing in Web sites, shows that "more and more people prefer using original content instead of the mobile version, especially for users of new generation mobile devices like iPhone and Android phones". Another result of this work was the users prefer to use the standard Web site instead of tailored mobile versions of Web site. But Schmiedl, Seidl and Temper [8] have a different opinion; in their research they conclude most of the users still prefer tailored versions. Kaikkonen [9] shows that the standard and the tailored Web sites are both used but for slightly reasons.

Considering only the e-Learning environment context, we agree to Maurer *et al.* [7] when they argue the user prefer to use the standard version instead of mobile version of Web site, and this is one of the motivations of our work to study the third kind of solution to access e-Learning environment using mobile devices.

Here, we studied the use of touchscreen devices into e-Learning environments and distinguish the problems into categories. Shrestha [6] studied the use of joystick and a small screen to browsing into some Web sites, while Maurer *et al.* [7] considers touchscreen devices. All these researches point out some interaction problems, but do not classify them if they happen due the platform changing or due the interaction changing.

III. TELEDUC E-LEARNING ENVIRONMENT

The TelEduc is a teaching and learning environment developed by the Nucleus of Applied Informatics in Education (NIED) and the Institute of Computing (IC), State University of Campinas (UNICAMP), and adopted in several public and private institutions, like UNICAMP through Ensino Aberto project [10].

The TelEduc environment was conceited in the end of 90, born with the Cerceau´s Master dissertation (1998), with professor Heloísa Vieira da Rocha as advisor, applying constructivism theory [10][12] in situated learning [13] or in contextualized learning [14] for teacher´s continuance formation. In 2001 February, the first free version was released over GNU General Public License (GPL), an unprecedented fact in the Brazilian Educational Software scenario. Many public and private institutions adopted the TelEduc as platform, increasing the TelEduc user's community, and consequently, the development demand. This fact culminated in the release of TelEduc version 3.0 in March 2002. The version 3.0 was completely redesigned and optimized, reason for TelEduc project was awarded by ABED (Brazilian Association for Distance Education) in the "Research and Development about Distance Learning" category. In August 2011, TelEduc version 4.3 was released, with its user interface redesigned to improve user tasks and be more similar than popular Web sites.

TelEduc is a system that aggregate administration, management and communication tools designed to support teaching and learning activities. Some tools allow users to create content, other ones allow synchronous or asynchronous communication among users, and manage participants and courses. The course page of TelEduc 4.3 is structured in two parts: the left one (Fig. 1a) has a list of all tools available and in the right one (Fig. 1b) the content of the selected tool.

In the course showed at Fig. 1a, the teacher dispose the Course Dynamic, Agenda, Readings, Support Material, Activities, Chat, Mail, Discussion Forum, Frequently Asked Questions, Portfolio, Groups and other tools available. Fig. 1b shows the user interface to visualize a Support Material item, where the teacher can change the title, content, attach or remove files or links and see and write comments, and the student can see the item, download the attached files and

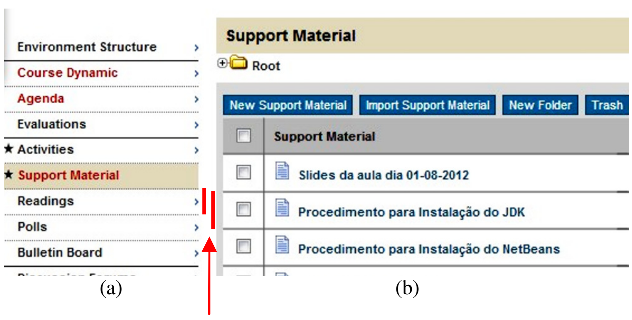


Figure 1. Height difference between menu items (a) and Support Material items (b).

visit the posted links.

Agenda is the ongoing course home page and shows the course´s program for a given period (daily, weekly, etc.). Agenda is an important tool because organize the activities that must be done in a specific period, similar teachers do in the beginning of a presence class.

The Support Material is a tool that provides an area for file storage and sharing among course participants, named Support Material Area. To store an item in this area, the user needs to be a coordinator or an instructor. When the user stores an item in support material area, she can specify the sharing type: i) not shared; ii) shared only with users who have instructor role; or iii) shared with all participants. Users with student role can access the stored items published with all participant sharing type, read their content, visit their links and download their attached files.

The Readings, Activities and Frequently Asked Questions tools have similar features and user interface with Support Material tool, but different purposes. Readings tool is used to publish relevant documents, like books, magazines, news and articles. The Activities tool is an area to publish activities to the accomplished during the course, like home work descriptions. The Frequently Asked Questions tool contains a list of the most frequently questions done by the participants during the course and their respective answers.

Tools like Discussion Forums and Mail are used to participant communication, supporting text message exchange in asynchronous mode. To synchronous communication, there is the Chat tool, its features is similar to Web chat sites.

The Portfolio is a communication tool that aims to promote the collaboration among participants through the sharing of "items" (documents, presentations, programs, links, etc.). So the Portfolio tool provides an area to item storage and sharing for each participant (user or group of users) within a course.

The Bulletin Board tool is a dedicated space where all the participants can post information considered relevant to the course content.

The Agenda, Activities, Support Material, Readings, Bulletin Board, Discussion Forums, Mail and Portfolio allow users to create text content using a text editor, the CKEditor [15]. CKEditor is a third-party WYSIWYG text editor to be used inside Web pages, bringing to the Web application common editing features found on desktop editing text applications, but CKEditor version 3.3 does not work on mobile devices. In [16], we studied some problems in use of third-party software on TelEduc to create and to visualize documents, describing that mobile compatibility needs to be considered to not prejudice mobile users.

The Agenda, Activities, Support Materials, Readings, Mail and Portfolio tools allow attaching files into them items in similar way: the user clicks in the "Attach File" button, select the file to be attached using a dialog and, after the click on the save button, the file is uploaded.

To provide content, TelEduc uses the Web infrastructure, more specifically, hypertext with images, links, audios and videos. All these media can be published as content in tools like Agenda, Support Material and Readings.

Since the e-Learning environments need to be easy to use for users with different levels of Web experience, the usability is an important nonfunctional requirement. TelEduc was designed in an iterative design-evaluation process to have good usability and the user interface does not impair teaching and learning activities, so many usability evaluations were done. The accessibility is another nonfunctional requirement desired for TelEduc, to allow impaired people to use the environment without meet barriers or obstacles.

TelEduc was designed to use a mouse and keyboard as input devices, and a medium screen size as output device. TelEduc is better visualized into 1024x200 pixels screen resolution. Visualize it into a lower screen resolution cause some user interface problems like dispose interface components in wrong position.

Since the e-Learning environments are available on the Internet, this software can be accessed by smartphones and tablets nowadays and the developers need to study how to allow all features into these devices. Access the environment in anywhere and anytime is one of the biggest attractions, but research is necessary to have a user interfaces with high usability and good user experience. To reach it, it is necessary identify cross-modality problems when TelEduc is used in touchscreen devices.

IV. MATERIALS AND METHOD

To investigate which problems happen when the users use mobile devices to browsing a Web application, we adopted the following method: a human-computer interaction (IHC) specialist, using a touchscreen device, accesses the Web application and collect interaction problems. To each collected problem, the specialist checks which ones from the three devices the problem happens and so classified into the categories: platform and modality-independent problem, cross-platform problem and cross-modality problem.

We used a Motorola Milestone smartphone [17] and a tablet PC [18]. The Motorola Milestone has a 3.7 inches multitouch display with 133MB internal storage memory expansive up to 32 GB with a memory card, 600 MHz Cortex-A8 processor and 256 MB RAM, a 5MP camera, GPS and wireless connection by Wi-Fi 802.11 b/g and bluetooth. Android 4.0.3 [19] was used as operation system (OS). Motorola Milestone has a proximity sensor, an ambient light sensor, a 3-axis accelerometer and a geomagnetism sensor to provide orientation with respect to Earth´s magnetic field. To browse in the Web application, the specialist uses the Android stock Web browser. The proximity sensor and the accelerometer can be used to interact with applications, but the stock Web browser does not use these features as an input device. Only the touchscreen was used as input device.

The Tablet PC is a computing device designed to "imitate" a notebook, allowing the user interact with a pen. Resuming, the Tablet PC has the following hardware characteristics: (i) Pen sensitive screen; (ii) Screen that

 ISBN: 978-1-61208-250-9

allows different positions; (iii) Wireless network access by WLAN and bluetooth technology; (iv) Microphones and embedded loudspeakers; (v) Keyboard (some models the keyboard are detachable); (vi) Batteries.

In this exploratory study, we used a HP TouchSmart TX2-1040br, a 2.2 GHz dual-core processor computer with 3 GB RAM and a 12” touchscreen. The installed operation system was Windows Vista with Portuguese manuscript recognizer. This model has the design similar to HP laptops but it is equipped with the described hardware for Tablet PC. To the study case, the specialist only used touch in the interaction; she did not use the Tablet PC´s keyboard or pen. Chrome browser version 22.0.1229.94 m [20] was used to navigate through Web application.

To classify the problem, the interactions in the three devices were compared (Fig. 2). If the problem appears in all devices, the probably of the problem be a platform and modality- independent problem is high. If the problem only happens in touchscreen devices, the probably of the problem be a cross-modality problem is high. But if the problem only happens in smartphone or only in tablet, maybe it is a cross-platform problem. So, to distinguish cross-platform problem, cross-modality problem and platform and modality-independent problem we needed to use these three devices. The Tablet PC and Desktop have the same platform (Widows Vista and Chrome). Tablet PC and the smartphone are both touchscreen devices and different platform (smartphone uses Android).

V. Interaction Problems

The specialist, using both touchscreen devices, found some interaction problems. An overview of the main problems is presented in this section.

A. *Problem 1: Fat Finger Problem*

At TelEduc, the tools available to be used in a course are listed in a menu disposed in the left side, and each tool is an option in this menu (Fig. 1a). TelEduc uses vertical lists not only for the tools menu, it uses vertical list to shows items to choose, e.g., in Support Material tool, each item is an option in a vertical list (Fig. 1b). An option in items lists (36 px) is higher than an option into tools menu list (23 px). This is a small difference, but the specialist points out problem to select a tool into the menu. Due to the small option menu height the specialist had problem to select a tool, triggered

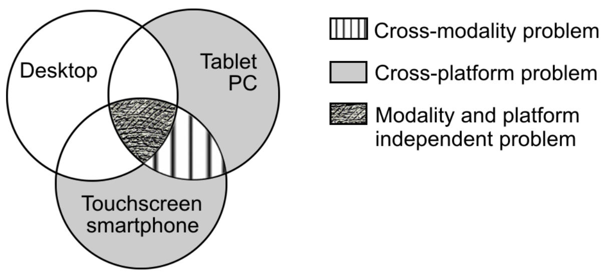


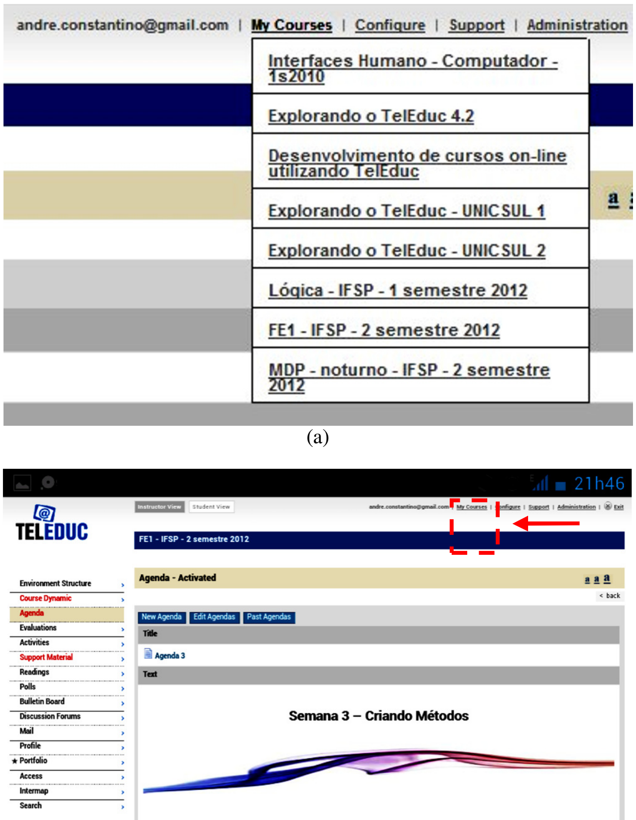
Figure 2. Categories of identified interaction problems on used devices.

the upper or down one, but this problem happen because the touched area is bigger than the clicked area when the user uses a mouse. This problem is called fat finger problem [21], when a user tries to acquire a target, the center of the contact area tends to be located a couple of millimeters off the target location—typically “below” the target [22].

The observed fat finger problem happens in both touchscreen devices, but more into smartphone probably due to page adaptation to show all content in the screen. The specialist observed the problem happens more into menu selection, because the space among menu items are not enough too large. The specialist observed the problem did not happen when selecting an item in the Portfolio or Support Material tool, probably because the distance among the items.

B. *Problem 2: Mouseover functionality*

TelEduc have My Course link to see the enrollment courses the user are and go to the Courses page. In desktop, the user can see the enrollment courses putting the mouse pointer over My Courses option, localized in screen right-top (Fig. 3a, the courses name is in Portuguese). If the user clicks in My Course link, she goes to Courses page. Touch screen devices does not show this menu since there is no *mouseover* action (Fig. 3b, the Agenda content is in Portuguese), and when the user try to see the enrollment courses, she triggers



(a)

(b)

Figure 3. Position of the My Courses menu (a) on Desktop and (b) on smartphone, the user cannot trigger the menu.

the My Course link and goes to the Courses pages.

This problem happens due the JavaScript *mouseover* function. It is common in Web applications the use of JavaScript to improve their user interfaces. But, some features of the JavaScript can cause interaction problems, like the *mouseover* function. The Android stock Web browser on smartphones makes a map between user´s action to browsing event, but there is no valid mapping to activate the *mouseover* function, since the used devices do not identify finger proximity, so it is not possible to have a feature similar to *mouseover* to do with the finger. The same problem happened on tablet.

Some of features in Table I are triggered by gesture. Using gesture into touchscreen devices the user can go forward, backward, scroll up and scroll down the page.

C. Problem 3: Gestures

Motorola Milestone and HP Tablet PC allow user interaction using gestures. To browse using these devices, the user can use one or two fingers and make gestures. Default gestures are to zooming; scroll up and scroll down a page; forward and backward pages; and select, copy and paste text. For novice users, it is not clear how to make the gesture. This is a discoverability and visibility problem as Nielsen and Norman related [23].

But, the user does not add specific gesture to use with a Web application, and the Web application cannot add specific gestures to user browser. This limitation does not allow TelEduc have gestures to create a new portfolio item, to select some Mail message or to delete a Support Material item. Since gestures have the promise to brings a powerful interaction [24], the Web application does not get all the promised potentiality, and the users gesture are limited to scroll up or scroll down a page; forward or backward pages; zooming; and select, copy and paste text.

TABLE I. MAPPING BETWEEN USER ACTIONS AND BROWSING EVENT IN THE ANDROID STOCK WEB BROWSER

Browsing Events	**Using mouse**	**Using finger**
Link activation	Left button click	Touch with one finger
Menu drop down	Right click button	Touch with one finger and hold
Scrolling text	Mouse click over scroll component interface	Touch with one finger and drag
Zooming	Not possible (needs change browser configuration)	Touch with two fingers and spread/pinch or double tap
Select text	Left mouse click over the text beginning and drag until the text end	Touch with one finger over the text and hold, release and drag the selection text component
Copy text	Select the text, click on right mouse button and choose copy	Select the text, click the option button and choose copy
Paste text	Right mouse button and choose paste	Touch with one finger and hold, choose paste option

D. Problem 4: Device features

Android platform specification defines four physical buttons: Back, Menu, Home and Search. Android platform allows developers customize the reaction of these standard buttons, like use search to find in application data or show the application menu instead of the default menu. This is one of the differences between Web applications and Android applications. Web applications do not have this possibility, and the action buttons are defined by the Web browser. So the search button, instead of searching into Web application data, opens the URL field (search a page).

Smartphones have a lot of features and the Web application cannot use. The specialist tried to post a photo and a video in her Portfolio, but the browser does not upload them and does not show a message error.

E. Problem 5: Third-party text editor

TelEduc uses CKEditor to allow users write rich text instead of simple plain text (Fig. 4a). But CKEditor does not work in Android devices (Fig. 4b), but works in the tablet.

F. Problem 6: menu activation

TelEduc menu item is only activated when the user click in the menu item text, a little different from the computers menu interaction. This difference is easy to understand when browsing in the menu, since the mouse pointer does not change when the pointer is over the menu item. In the case of smartphone and tablet, the user can try many times touching the space in front of the text to understand there she must tap the text.

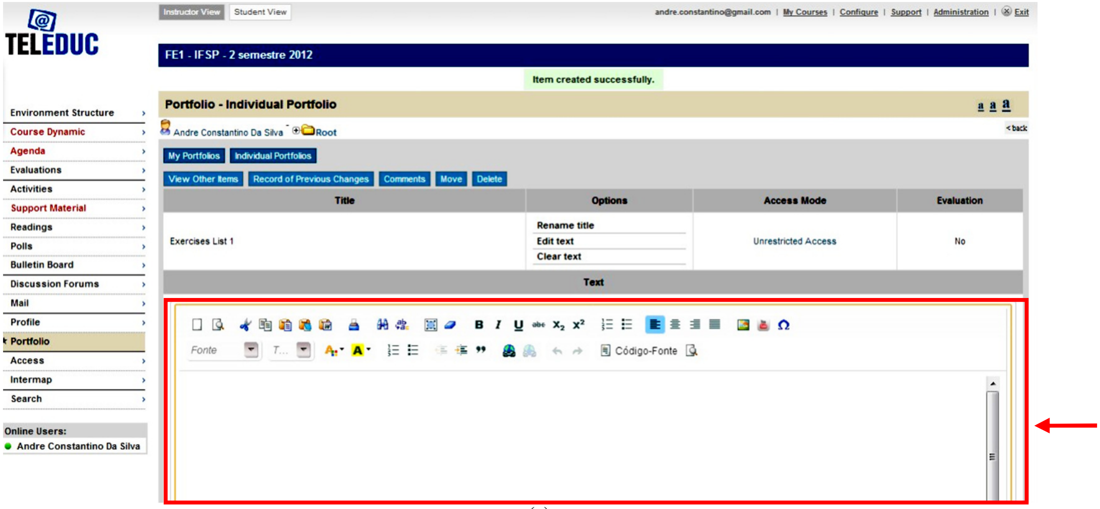
Table II summarizes the identified problems and the platform where they happen. The next section discusses and classifies these problems.

VI. CROSS-PLATFORM AND CROSS-MODALITY PROBLEMS

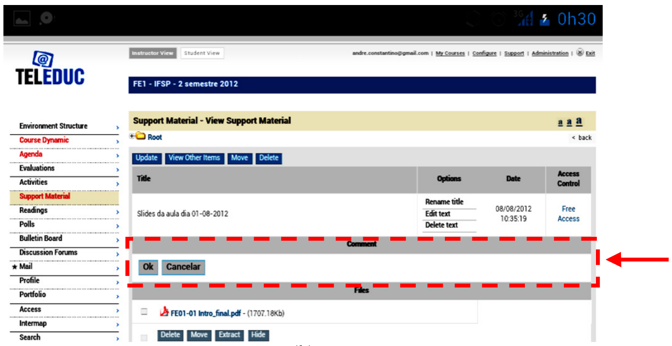
The fat finger problem (Problem 1), *mouseover* functionality (Problem 2), gesture integration (Problem 3) and menu activation (Problem 6) happened in both smartphone and tablet. Gesture integration can be classified

TABLE II. IDENTIFIED PROBLEMS AND THEIR CLASSIFICATION

	Problem	**Touchscreen smartphone**	**Tablet PC**	**Desktop**	**Classification**
1	Fat finger problem	yes	yes	no	cross-modality
2	*Mouseover* functionality	yes	yes	no	cross-modality
3	Gesture integration	yes	yes	no	cross-platform
4	Device features gap	yes	no	no	cross-platform
5	Third-party text editor	yes	no	no	cross-platform
6	Menu activation	yes	yes	yes	platform and modality-independent



(a)



(b)

Figure 4. Position of the rich text editor in TelEduc a) the editor is displayed on Desktop and tablet PC and b) it is not displayed on Android-based devices.

as platform problem, though it happens on both devices, this problem is not directly related with touch; this is a lack of functionality. Fat finger problem and the use of *mouseover* functionality were considered as cross-modality problems. They happened when the user interface, designed to be used with an interaction style, is accessed by different interaction styles. If the application was designed to be used with touchscreens, the designer would choose the suitable interface components size and the spacing among them, decreasing the occurrence of the fat finger problem. In the case of *mouseover* functionality, the designer would choose a better way to show the menu, e.g., when the My Course link is activated, the menu is showed instead of go to Courses page.

Menu activation (Problem 6) was considered as platform and modality-independent problem, since the problem happened in desktop computers too, but when the user is using a smartphone the problem turns more severe and be easier to find.

Device features gap (Problem 4) only happens in smartphone due the difference between the Android platform and Desktop Operating Systems. This problem was classified as cross-platform problem.

The problem with no display the CKEditor (Problem 5) was classified as cross-platform problem. CKEditor works fine in the tablet, though the specialist has difficulty using a text editor.

VII. CONCLUSION AND FUTURE WORK

Almost Web sites and applications are developed thinking to be used with mouse and keyboard as input device, and medium size screen as output devices. But smart phones with touchscreen have enough computational power to access Internet, so these devices allow users browse into Web applications by touching. When a smartphone is used to access Web applications we can have modality and/or platform changing.

Changing the interaction style or platform brings interaction problems, impacting on usability. Which problems, if we consider the software have good usability in desktop computers, appears when we used a non-specified interaction style to browse a Web application? And which problems are not related with the platform changing, just only with the modality changing? We call cross-modality interaction problems. We developed this work trying to identify cross-platform and cross-modality interaction problems using TelEduc, an e-Learning environment developed to be used with mouse and keyboard as input devices, in touchscreen devices.

To identify problems, a HCI specialist analyzed the interaction and observed problems, classified them into platform and modality-independent problems, cross-platform problems and cross-modality problems. Through this work we prove software designed to be used with some interaction styles may have problems to be used with other interaction style. In this paper, we presented 6 problems to distinguish cross-platform and cross-modality problems, and show interaction problems that are more severe when the interaction style is changed.

It is important highlight we used devices with limitation, e.g., the algorithm who decides which user interface component the user touched. The algorithm accuracy may prejudice the interaction problem identification. Disregarding these limitations, it is clear the need to a better integration between platform and Web applications to increase the user experience, gestures need be more explored when user browsing.

Another important result is the perception of the changing interaction styles allows highlight existing usability problems.

As future works, we planned to study solution for these problems, the relationship between the problem and the code and study cross-modality problems for other input devices, like pen.

ACKNOWLEDGMENT

The authors thank the CAPES and CNPq for financial support and for FAPESP through TIDIA-Ae Project (n. 05/60572-1), which provided the equipment.

REFERENCES

[1] Moodle Trust. Moodle.org: open-source community-based tools for learning. Available at <http://moodle.org>. Accessed in January 2013.

[2] SAKAI Environment. Sakai Project | collaboration and learning - for educators by educators. Available at <http://sakaiproject.org>. Accessed in January 2013.

[3] TelEduc Environment. TelEduc Ensino à Distância. Available at <http://www.teleduc.org.br>. Accessed in October 2012.

[4] Moodle Trust. Moodle App. Available at: <http://docs.moodle .org/dev/Mobile_app>. Accessed in January 2013.

[5] M. J. Casany, M. Alier, E. Mayol, J. Piguillem, N. Galanis, F. J. García-Peñalvo, and M. Á. Conde. "Extending Moodle Services to Mobile Devices: The Moodbile Project". Proc. of the 6th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2012), IARIA, 2012, pp. 24-28.

[6] S. Shrestha. "Mobile Web Browsing: Usability Study". Proc. of the 4th International Conference on Mobility Technology, Applications and Systems (Mobility 2012), ACM Press, 2007, pp. 187-194, doi: 10.1145/1378063.1378094.

[7] M.E. Maurer, D. Hausen, A. De Luca, and H. Hussman. "Mobile or Desktop Website? Website Usage on Multitouch Devices". Proc. of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI 2010), ACM Press, 2007, pp. 739-742, doi: 10.1145/1868914.1869018.

[8] G. Schmiedl, M. Seidl, and K. Temper. "Mobile Phone Web Browsing –A Study on Usage and Usability of the Mobile Web". Proc. of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI´09), ACM Press, 2009, doi: 10.1145/1613858.1613942.

[9] A. Kaikkonen. "Full or tailored mobile web – where and how do people browse on their mobile?". Proc. of the International Conference on Mobile Technology, Applications and Systems (Mobility 2008), ACM Press, 2008, pp. 111-111, doi: 10.1145/1506270.1506307.

[10] M. A. Franco, L. M. Cordeiro, and R. A. F. del Castillo. "O ambiente virtual de aprendizagem e sua incorporação na Unicamp", Educação e Pesquisa, vol. 29, issue 2, Jun-Jul 2003, pp. 341-353.

[11] S. Papert, "Constructionism: a new opportunity for elementary science education". Proposal to The National Science Foundation. Massachusetts: Cambridge, 1986.

[12] J. A. Valente, "Computadores e conhecimento: repensando a educação". Campinas, Brazil: Gráfica da UNICAMP., 1993.

[13] J. Lave and E. Wenger, "Situated Learning – Legitimate peripheral participation". Cambrigde: Cambrigde University Press, 1991.

[14] J. A. Valente, "Formação de Professores: Diferentes Abordagens Pedagógicas". In O Computador na Sociedade do Conhecimento, J. A. Valente, Ed. Campinas, Brazil: UNICAMP/NIED, 1999, pp. 131-156.

[15] CKSource - Frederico Knabben. CKEditor - WYSIWYG Text and HTML Editor for the Web. Available at <http://www.ckeditor.com>. Accessed in January 2013.

[16] A. C. da Silva, H. V. da Rocha. "Improving e-Learning Environment With GPL Software - The visualization content case for computers and mobile devices". Proc. of the 6th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2012), IARIA, 2012, pp. 222-227.

[17] Motorola Mobility Inc., 'Milestone Android – Phone Without Compromise - Tech Specs - Motorola Mobility India', Available at: <http://www.motorola.com/Consumers/IN-EN/Consumer-Products-and-Services/MobilePhones/ci.Motorola-MI LESTONE-IN-EN.alt>. Accessed in January 2013.

[18] HP TouchSmart tx2-1040br Notebook PC (NL266LA) specification. 2009. Available at: <http://h10025.www1.hp.com/ewfrf/wc/product?cc=us&lc=en&dlc=en&tmp_geoLoc=true&product=3937930>. Accessed in January 2013.

[19] Google. 'Android'. Available at: <http://www.android.com/> Accessed in January 2013.

[20] Google Inc. Chrome Browser. Available at: <http://www.google.com/chrome>. Accessed in January 2013.

[21] D. Vogel, and P. Baudisch. "Shift: A Technique for Operating Pen-Based Interfaces Using Touch". Proc. of the 25th International Conference on Human Factors in Computing Systems (CHI'07), ACM Press, 2007, pp. 657–666, doi: 10.1145/1240624.1240727.

[22] C. Holz and P. Baudisch. "The Generalized Perceived Input Point Model and How to Double Touch Accuracy by Extracting Fingerprints". Proc. of the 28th International Conference on Human Factors in Computing Systems (CHI '10), ACM Press, Apr. 2010, pp. 581–590, doi: 10.1145/1753326.1753413.

[23] D. A. Norman and J. Nielsen. "Gestural Interfaces: A Step Backwards In Usability", Interactions, vol. 17, issue 5, Sep.-Oct. 2010, pp. 46-49, doi: 10.1145/1836216.18362287.

[24] D. A. Norman. "Natural User Interfaces Are Not Natural", Interactions, vol. 17, issue 3, May-Jun. 2010, pp. 6-10, doi: 10.1145/1744161.1744163.

 ISBN: 978-1-61208-250-9

Applying Commercial Digital Games to Promote Upper Extremity Movement Functions for Stroke Patients

Lan-Ling Huang[1] and Chang-Franw Lee[2]
[1, 2]Graduate School of Design
National Yunlin University of Science and Technology
Yunlin 64002, Taiwan
[1]lanlingh@gmail.com
[2]leecf@yuntech.edu.tw

Mei-Hsiang Chen[3]
[3]School of Occupational Therapy
Chung Shan Medical University/Chung Shan Medical University Hospital
Taichung 40201, Taiwan
[3]cmh@csmu.edu.tw

***Abstract*—The objective of this study is to evaluate the effectiveness, usability and satisfaction of conventional devices, Nintendo Wii and XaviX, on upper extremity rehabilitation patients in Taiwan, and to summarize a guideline for improvement design of such devices. Twelve stroke patients were divided in three groups: (1) Conventional, (2) Wii, (3) XaviX groups. Eight senior occupational therapists were interviewed about the usage problems and additional needs related to the use of these devices. The results show that Wii and XaviX could be equivalent to conventional rehabilitation devices for improving upper extremity motor functions. All patients in this clinical trial were satisfied with using the digital gaming devices for rehabilitation. The suggestions for improvement design in game devices are as follows. For the software interface devices: (a) To increase difficulty and the response time levels of the games need adjustment, (b) To record movement data and game scores each time, some device for recording is needed, and (c) The games need a Chinese version of the software interface. For the hardware design: (a) The hand controller must be interchangeable for the users, (b) The controller should be adjustable to fit different hand dimensions of the patients, (c) The game and controller movements need to be designed to correspond to real-life activities, and (d) The controller's operation needs to be simplified. These proposed guidelines would be necessary in order to embody design improvements of the devices.**

Keywords- effectiveness; usability assessment; commercial digital game devices; stroke; upper extremity rehabilitation

I. INTRODUCTION

The current clinical upper extremity rehabilitation devices are mostly static and provide no feedback to the patients [1]. Patients easily feel bored while repeating the same activity, hence generating a negative attitude toward the therapy process [2][3][4]. In order to increase mental satisfaction and physical vitality in rehabilitation therapy, some therapists have tried to use the existing commercial digital game devices (CDGD)in rehabilitation and have found effective treatment outcomes in addition to enhancing the patient's treatment motivation [3][5][6].

There are already many studies focused on digital game devices in rehabilitation such as Wii [2][3][5][6][8], Playstation EyeToy [5][8] and Kinect [4]. For example, a study examined the feasibility and safety of the Wii gaming system and compared it with recreational therapy in facilitating motor function of the upper extremity [7]. The results showed that Wii gaming technology represents a safe, feasible, and potentially effective way to facilitate rehabilitation therapy and promote motor recovery after stroke. However, some of Wii's disadvantages were that stroke patients found the control of the handset buttons difficult and frustrating to use; however, this obligatory hand use improved gross motor dexterity. Stroke patients did not simply play Wii sports, rather the device was used as a rehabilitation tool with targeted and movement goals aimed at reinforcing appropriate and coordinated motor patterns [6].

These existing devices are originally designed for entertainment with normal people with healthy physical and action conditions, not intended for rehabilitation therapy purposes or for people with physical disabilities [5]. Further confirmation and evaluation is necessary to see if a digital game device can really meet the user's usability needs in addition to its rehabilitation effectiveness. Usability testing is a method used to determine how easy a device is to use and to identify issues that must be addressed in order to improve the design and functionality of the device [9]. However, the usability of such devices for rehabilitation remains relatively rare [5].

Nintendo Wii and XaviX have been tested in clinical rehabilitation in several hospitals in Taiwan. However, whether the superiority of Virtual Reality systems can facilitate conventional therapy currently in use remains to be determined [2][10][11]. Further development is required to ensure that these devices are easy for patients and therapists to set up in a clinic and/or home setting [5]. Therefore, this study aimed to evaluate the effectiveness, usability and satisfaction of using Wii and XaviX in rehabilitation for stroke patients and for the occupational therapists, and to summarize a guideline for improvement design of such devices.

In the following sections, the methods applied and results obtained will be described respectively, and then implication of the results will be discussed, followed by a brief section of conclusion and future work.

II. METHODS

A pilot and double-blind clinical trial was implemented to evaluate the effectiveness between conventional devices, Wii, and XaviX in rehabilitation. In addition, the usability and satisfaction of using Wii and XaviX in rehabilitation were also assessed by stroke patients and occupational therapists.

 ISBN: 978-1-61208-250-9

A. *effectiveness evaluation*

1) Subjects. Stroke patients were recruited from the occupational therapy department of Chung Shan Medical University Hospital. Inclusion criteria were the following: (a) Hemiparetic with upper extremity dysfunction following a single unilateral stroke, (b) a history of first-time stroke (3-48 months post stroke, (c) the required upper extremity rehabilitation convalescent levels were Brunnstrom stage III to IV, i.e., having basic upper extremity synergies to perform joint movement voluntarily, (d) ability to communicate, (e) able to understand and follow instructions. Exclusion criteria were the following: (a) engaged in any other rehabilitation program during the study and (b) serious aphasia or cognitive impairment. Each patient gave informed consent. This study was approved by the Human Research Ethics Board of Chung Shan Medical University Hospital.

2) Settings. Conventional equipment, Wii, and XaviX were used in this trial (Table I). Each group was assigned to use two games or equipment in the additional treatment. The games and equipment for the groups were selected by three occupational therapists, and were considered as similar in training effect on upper extremity movements. For Nintendo Wii, two games (boxing and bowling) of Wii Sport were selected to use in this trial. As for XaviX, two games (bowling and ladder climbing) were selected to use in this trial. Ladder climbing contains three levels (easy, normal and difficult). While playing bowling, the user needed to hold a soft bowling ball, fixed to the hand with a safety belt. In ladder climbing, the user needed to wear glove sensors in the palms of both hands [13]. Corresponding to the Wii games (Bowling and Boxing) and the XaviX games (Bowling and Ladder climbing), two conventional equipment were selected in this trial. They are the Curamotion exerciser and the Climbing board and bar.

3) Functional assessments. Four functional assessments were used as follow:

a) Fugl-Meyer Assessment of Physical Performance (FMA) [14]. It was used to evaluate the motor functions. The upper extremity motor test part with a possible highest score of 66 was adopted in the evaluation. The reliability of Fugl-Meyer Assessment is generally considered reliable [15].

b) Box and Block Test of Manual Dexterity (BBT). It was used to test gross manual dexterity of a patient's affected side [16]. In the test, the patient was asked to move as many cubes (of side length of 2.5 cm) as possible using only the thumb and index fingers during a timed 60s trial.

c) The Functional Independence Measure (FIM). This scale assesses physical and cognitive disability. The scale includes 18 items, of which 13 items are physical domains based on the Barthel Index and 5 items are cognition items. Each item is scored from 1 to 7 based on level of independence, where 1 represents total dependence and 7 indicates complete independence. This measurement was assessed and shown to have high reliability and validity [17][18][19].

d) Upper extremity range-of-motion (ROM). This is used to assess the passive and active range-of-motion on the affected side [20].

4) Duration of intervention. The training comprised 20 sessions during 2 months, with each session lasting 30 minutes (excluding set-up time). The effectiveness was evaluated before and after completing the 20 training sessions. In addition to these trainings (Wii, XaviX, and Conventional) in this study, all patients also received at least 1 hour of occupational therapy and 1 hour of physical therapy.

B. *Usability and satisfaction evaluation.*

An interviewer-administered questionnaire was designed to evaluate usability and satisfaction of using Nintendo Wii, and XaviX. The questionnaire included two parts:

1) Usability. Three open-ended questions about device set-up, game-playing, and performance feedback were asked to the patients for them to answer orally. By these questions, we expected to understand the context of use problems encountered by the patients.

2) Satisfaction. Three 5-point Likert type questions about satisfaction, motivation, and fun were then presented on a sheet of paper for the patients to answer. By this trial, we expected to understand the patients' satisfaction with using the digital gaming devices in clinics.

C. *Usability evaluation by occupational therapists*

Expert interview was conducted to survey the current situation of the use of digital game devices and evaluate the usability of such devices in upper extremity rehabilitation.

1) Locations. Two hospitals were selected in a preliminary investigation: Taipei Veterans General Hospital (TVGH) and Kaohsiung Veterans General Hospital (KVGH). Wii is used in KVGH and Xavix is used in TVGH.

2) Subjects. As a professional, the occupational therapist possesses expertise about therapy theory and also has abundant experiences, and these are useful for the evaluation of rehabilitation devices or the commercial products applied in rehabilitation. Selection criteria for the interviewed therapists were as follows: at least 5 years work experience in occupational therapy, and at least one year experience in adopting the digital game device intervention in rehabilitation treatment.

3) Contents of the interview. Semi-structured interviews were used, which mainly consisted of two parts: (a) Therapist personal profile - gender, age, hospital name, and work experience in years, and (b) Questions about the usability evaluation of the Wii and XaviX.

D. *Observation of the usage of the CDGD in clinics.*

During field observation, the researchers observed the process of devices being operated independently by the patients or with help of the therapist. From the point of view of ergonomic design, the researcher would also take notes of usability issues of the devices.

E. *Data analysis.*

The characteristics of the study groups were described as mean and SD. Differences in baseline characteristics between the three groups were analyzed by One-Way ANOVA. Each group was analyzed using paired-samples t-

 ISBN: 978-1-61208-250-9

tests for pre- and post-therapy values. For each group, the effectiveness index of each functional assessment was calculated, that is, the post-test score minus the pre-test score, then divided by the maximum possible progress (possible highest score of the assessment minus pre-test score) [21]. To analyze the interview data from occupational therapists and stroke patients, the recording was firstly transcribed verbatim. Similar opinions were combined and all unique responses were independently itemized for further discussion.

III. RESULTS

A. *Effectiveness evaluation of stroke patients*

1) Characteristics of the stroke patients. Twelve consecutively screened stroke patients finally completed the trial, with a mean age of 50.42 years (SD=11.20). The characteristics of the patients in the three study groups are shown in Table II. There are no statistically significant differences between the three groups with regard to age, time from stroke onset, Brunnstrom stages of affected side, or FMA pre-treatment score.

2) Pre and post-gaming for all stroke patients. All groups showed statistically significant improvements on the FMA, BBT, FIM and 4 items of ROM assessment scales from pre-treatment to post-treatment (Table III).

3) Between-group differences in score changes for effectiveness index. The results show that effectiveness index of each group had improvements for arm functions in stroke patients on FMA, FIM and ROM, but there was no statistically significant improvement among the three groups (Table IV). For the FMA results, the Xavix group had a better effectiveness index score than the Conventional and Wii groups.

TABLE I. GAMES AND CONTROLLERS OF THREE GROUPS

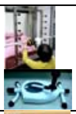
Group	Devices		Console	Controller	Program
Convention	Climbing board and bar Curamotion exerciser				Conventional therapy (60 min) + both the Climbing board and bar, and Curamotion exerciser (30 min. in total)
Wii	Bowling Boxing				Conventional therapy (60 min) + both the Bowling and Boxing games (30 min. in total)
XaviX	Bowling ladder climbing				Conventional therapy (60 min) + both the Bowling and Ladder climbing games (30 min. in total)

TABLE II. CHARACTERISTICS OF THE STROKE PATIENTS

characteristics	Convention	Wii	Xavix	p-value
number of subjects	4	4	4	
gender (male/ female)	4/0	4/0	3/1	
age (years) (mean, SD)	53.3 (13.3)	47.3 (13.9)	50.8 (8.1)	0.784
time from stroke onset to involve gaming (months) (mean, SD)	4.9 (1.7)	8.7 (3.7)	11.0 (12.9)	0.551
paretic side (left/ right)	2/2	1/3	3/1	
Brunnstrom stages (mean, SD)	4.0 (0.0)	4.8 (1.0)	4.0 (0.8)	0.291
(proximal/distal)	3.3 (1.0)	4.8 (1.0)	3.0 (1.8)	0.181
FMA (pre-gaming) (mean, SD)	28.8 (3.6)	49.8 (17.1)	36.5 (19.9)	0.201

TABLE III. PRE AND POST-GAMING FOR ALL STROKE PATIENTS

Assessments	Pre-gaming mean (SD)	Post-gaming mean (SD)	p-value
FMA (UE)	38.33 (16.54)	44.42 (16.52)	0.000*
BBT	15.83 (18.91)	20.50 (21.63)	0.054*
FIM	108.50 (14.99)	112.08 (13.31)	0.005*
ROM - shoulder flexion	114.25 (59.70)	136.25 (53.52)	0.148
shoulder hyperextension	34.17 (17.30)	39.58 (9.88)	0.090
shoulder adduction	83.33 (34.47)	107.92 (24.81)	0.002*
shoulder abduction	106.67 (47.31)	131.25 (48.58)	0.016*
shoulder internal rotation	42.92 (25.18)	47.75 (28.04)	0.263
shoulder external rotation	42.92 (29.58)	52.92 (32.92)	0.124
elbow flexion	111.67 (30.25)	125.00 (21.64)	0.112
forearm supination	35.00 (34.97)	52.92 (36.83)	0.010*
forearm pronation	45.00 (41.40)	52.08 (46.00)	0.068
wrist flexion	31.92 (35.09)	42.50 (33.06)	0.078
wrist extension	22.75 (26.18)	30.00 (27.63)	0.065
wrist ulnar deviation	12.08 (16.71)	15.42 (17.12)	0.087
wrist radial deviation	6.25 (8.56)	10.67 (10.17)	0.028*

*p value for pre-gaming versus post-gaming

TABLE IV. BETWEEN-GROUP DIFFERENCES IN CHANGE SCORES

Groups	Convention	Wii	Xavix	p-value
	EFFECTIVENESS INDEX, MEAN (SD)			
FMA (UE)	0.21 (0.50)	0.19 (0.15)	0.34 (0.46)	0.725
FIM	0.24 (0.13)	0.11 (0.15)	0.17 (0.21)	0.559
ROM - shoulder flexion	0.33 (0.58)	0.33 (0.47)	0.36 (0.43)	0.996
shoulder hyperextension	0.68 (0.22)	0.25 (0.50)	0.06 (0.13)	0.063
shoulder adduction	0.44 (0.23)	0.46 (0.53)	0.35 (0.27)	0.900
shoulder abduction	0.33 (0.42)	0.35 (0.40)	0.40 (0.41)	0.971
shoulder internal rotation	0.02 (0.16)	0.08 (0.03)	0.01 (0.08)	0.628
shoulder external rotation	0.06 (0.45)	0.58 (0.50)	0.27 (0.29)	0.262
elbow flexion	0.07 (1.17)	0.40 (0.49)	0.14 (0.34)	0.811
forearm supination	0.25 (0.17)	0.69 (0.47)	0.38 (0.48)	0.317
forearm pronation	0.47 (0.55)	0.50 (0.58)	0.25 (0.50)	0.780
wrist flexion	0.22 (0.27)	0.43 (0.42)	0.50 (1.00)	0.817
wrist extension	0.09 (0.20)	-0.78 (2.16)	0.34 (0.43)	0.456
wrist ulnar deviation	0.06 (0.07)	0.17 (0.33)	0.11 (0.22)	0.828
wrist radial deviation	0.27 (0.36)	0.90 (1.05)	0.19 (0.38)	0.313

B. *Usability and satisfaction evaluation.*

1) Characteristics of the interviewed occupational therapists. Eight therapists were selected and interviewed, three (one male [A] and two females [B, C]) were from KVGH and five (two males [D, E] and three females [F, G, H]) were from TVGH. They had an average age of 35.1 yrs (SD=6.3) and work experience of 10.8 years (SD=6.3).

2) Usability assessment by occupational therapists.

T1.Effectiveness. All eight therapists agreed that the Wii and XaviX are effective in upper extremity rehabilitation, and could enhance the patient's treatment motivation and pleasure. Therapist D commented that some of the game projects with their controllers can effectively strengthen the training of reaching and grasping movements and achieve better therapy effectiveness as compared to the traditional equipment. In addition, therapists mentioned that the existing conventional equipment are still important in rehabilitation and not possible to be fully replaced by digital gaming devices. The traditional equipment do not require the patient to react in time and restrict his/her movement speed, hence the patient can repeat the movement as many times as he/she wishes and at a speed at his/her own control. Therefore, the digital game devices can play a supporting role in rehabilitation treatment by providing diversification and interesting game projects for improvement of patient motivation toward the treatment.

T2. Ease of Use. Three therapists (A, B, C) considered that Wii is easy to set up, except some errors may occur when setting up the game software items. The usage problems are the following: (a) The current software interface is in Japanese and not easy to understand, hence it is prone to cause errors in the set-up process; (b) The required response time of the game is too fast and not easy to keep up with for the patients; (c) Some patients may have difficulty to hold the hand controller, hence the need for additional bandages to tie it on the hand; (d) For some patients, the games are too difficult.

For XaviX, five therapists agreed that the hardware is easy to set up, except the software interface operation may cause occasional mistakes. The usage problems are the following: (a) The current software interface is in Japanese and not easy to understand, hence it is prone to cause errors in the set-up process, especially for the games of visual perception and memory training, which are impossible to operate without literacy of the Japanese language; (b) The sensor is not sensitive enough, e.g., the action may obstruct the reflective film and interrupt the detection by the sensor; (c) The controller gloves have only one single size, thus, they do not necessarily fit to all patients; (d) Although the games are available in three levels of difficulty, the differences between the difficulties are too abrupt and hard to meet the patient's required degree of difficulty.

T3. Comfort. The results show that, for both Wii and XaviX, none of the eight therapists had received any complaints from the patients concerning discomfort in operating the controllers. However, therapist A commented that the controller of Wii often fell off from some patients' hands even with the safety ring held around the wrist. Therefore, he proposed that the controller be designed to be adjustable in size or to have a banding strap to help the patient to hold it with. Some patients cannot press the buttons on the controller to operate the menu due to incomplete recovery of the hand functions, hence they need to be assisted by the therapist for such operations before they can proceed with the game. Therefore, it is desirable to redesign the interaction mode between the controller and the game. For XaviX, therapist D suggested that this device provides a different set of controllers, which are suitable for different forms of motor training for the patients. For instance, the use of gloves, bowling balls, and other controllers can be useful for training the hand grip function.

T4. Acceptability. All eight therapists agreed that most patients can accept the use of Wii or XaviX for treatment. Therapist B explained that people interested in the game have a very high degree of acceptance of such devices. However, some elderly patients tend to prefer using the traditional rehabilitation equipment, because they feel that the traditional equipment are some physical objects that can be held and manipulated, hence giving a feeling of the effect of treatment. On the contrary, digital game devices give an impression of laxness and frivolity, and hence are inefficient and ineffective for those patients.

3) Usability assessment by stroke patients

S1. Device set-up. Two patients of the Wii group mentioned that they could set up the device and then operate the Remote to choose a game from the main menu to play by themselves. However, the other two patients needed the therapist to assist them to set up the device, and to operate the Remote to choose a game from the main menu for them to play. One patient of the XaviX group mentioned that he could set up the device and then operate the controller to choose a level of the game from the main menu to play by himself. The other three patients mentioned that they could only do some of the steps, i.e. to choose a level of the game from the main menu to play.

S2. Game-playing. Four patients that participated in the Wii group mentioned that they encountered only one usage problem: 'during the bowling game, they felt that it was difficult to press the control pad and to hold down the B button while using the device. Therefore, they usually had to use the unaffected hand to help. This difficulty is understandable, because their fingers, still under restoration, were weak and clumsy, hence they would need more time to operate the buttons. The four patients that participated in the XaviX group mentioned their respective usage problems as follows: (1) the safety belt of the bowling ball loosened easily, (2) the weight of the soft bowling ball was too light to feel: "I felt only the return swing of the hand. It didn't seem to be effective in training the strength of the upper extremity muscles." (3) Could not play well: "In the first two times I was not sure if my posture was correct, so I felt somewhat frustrated." (4) The bowling ball could not be gripped in the normal way: "My hand movement function still needs to be restored and the fingers curled, so the only way for me to play with the bowling ball was to hook the safety belt in my hand. In this way, the ball was unsteady when I swung my hand."

S3. Performance feedback. All eight patients agreed that they would pay attention to the scores of the games after playing. One of them mentioned that in order to upgrade the score, he would try to improve the strength of his upper extremity by practicing. The other two patients also mentioned that they would challenge the last score.

S4. Satisfaction. All the eight patients agreed that they were satisfied with using these devices for treatment (Wii: mean=4.0, SD=0.0; XaviX: mean=4.3, SD=0.5).

S5. Motivation. Eight patients agreed that these devices used in the treatment promoted their treatment motivation (Wii: mean=4.3, SD=0.5; XaviX: mean=4.0, SD=0.8).

S6. Fun. Eight patients agreed that these devices were more interesting than traditional rehabilitation equipment (Wii: mean=4.3, SD=0.5; XaviX: mean=4.0, SD=0.8).

Five of them mentioned that these devices provided useful information, such as the scores, audio and video feedback, as well as interesting interactions, making them feel positively toward doing the treatment activities.

IV. DISCUSSIONS.

Many studies have reported that digital gaming is able to promote motor recovery after stroke, but most trials were small in scale and had some design limitations [2][10][11]. However, one clear advantage of the use of these games for rehabilitation is the psychological effects. It facilitates interest and motor learning, promoting motivation through the in-built commentary prizes, visual and verbal cues, tapping into patients' innate competitive natures and into their desire for interaction [2][6]. These effects are unavailable from conventional rehabilitation equipment. These advantages may be a reason to affect the effectiveness of the patients' rehabilitation therapy. It is worth preserving and incorporating them into digital rehabilitation device design or conventional rehabilitation equipment redesign.

Effectiveness of Wii and XaviX could compare to conventional equipment. In order to be fairer when comparing the effectiveness between the three groups, we tried to reduce the differences according to the age range of the patients in each group. Although, no statistically significant differences in the effectiveness indexes between the three groups were found, patients in the Wii and XaviX groups had improved upper extremity motor functional ability. This result seems to demonstrate that Wii and XaviX could be equivalent to conventional rehabilitation equipment for improving upper extremity motor functions.

Therapists and patients' perspectives about satisfaction. Therapists considered that patients generally were willing to use the digital game for rehabilitation. Therapists mentioned that only some elderly patients tended to prefer using the traditional rehabilitation equipment, because they felt that the traditional equipment are tangible objects that can be held and manipulated, hence giving a feeling of their effectiveness in treatment. However, all patients in this clinical trial were satisfied with using the digital gaming devices for rehabilitation. They expressed the hope that the digital gaming devices could be included in conventional rehabilitation therapy programs, and they felt that they increased the diversity of the rehabilitation activities. In addition, they also suggested that it would be better if the controllers of such devices were adjustable in weight.

The interface was difficult to use. In this study, the stroke patients' upper extremity rehabilitation convalescent levels were in the Brunnstrom stages III to IV, their fingers were mostly not restored at all. Hence, they had difficulties in performing fine movements, such as pressing the buttons on the Remote. To resolve this usage problem, it seems that the existing XaviX-Eyehand operation could be applied for such patients. This software is operated by the Glove Sensors worn on the user's hands. The user can move a cursor with the movement of the gloved hand, and select a menu item by clenching his/her fist (to hide the reflective sheets on the Glove Sensors). Though such an operation is easier than operating a controller to press the button, it is not preferable for patients in these stages, who still have difficulty in clenching their fists. An improvement of the design can be achieved by having users only to move the cursor to a fixed-point for some seconds, indicating a purposeful choice. In addition, some newly developed technologies can also be adopted.

The problem about device set-up. Observation of the therapists and patients in setting-up the XaviX revealed that they spent much time in determining the sensing range between the host and the controller in order to optimize the operating area for the players. To improve this problem, it is recommended that the device be designed with a mechanism to sense the region in a graphic fashion. This mechanism would let users immediately know whether the controller is in the sensing region or not, hence enhancing the efficiency of device set-up.

The difficulty levels of the games were too limited. People with stroke have different levels and complex impairments of varying severity. The impairments may include weakness of the arm, impaired vision, a variety of cognitive and executive problems or a combination of these impairments. These all can reduce their upper extremity function and response capacities. However, the difficulty levels provided by most of the commercial digital game devices are unable to meet the real needs of clinical rehabilitation therapy. For example, the XaviX- ladder climbing has only three difficulty levels, which is clearly not enough. Furthermore, Wii Sports even has no grade distinction at all. It was observed that two of the patients in the XaviX group were slow in response and had difficulties to timely respond to the game task. Therefore, in order to meet the treatment needs of different upper extremity function levels of the stroke patients, the device should be designed with more difficulty levels. How to define the scope of the game difficulty levels requires further trials in the future.

V. CONCLUSION AND FUTURE WORK

From the results obtained from the interviews and observations, it was found that the digital game devices currently used in clinical rehabilitation have been evaluated by professional therapists as useful and effective in supporting the treatment, though some design improvements may be necessary. This report mainly presents the results of the interview with occupational therapists and patients and observations of the clinical use of existing digital game devices in rehabilitation. Conclusions can be summarized into the following points:

1. All patients in this clinical trial were satisfied with using the digital gaming devices for rehabilitation. In addition, therapists also suggested that it would be better if the controllers of such devices were adjustable in weight.

2. Design guidelines concerning the improvement of existing digital game devices can be synthesized as follows, where items *a* to *d* are about software design, and items *e* to *i* about hardware: a) To increase the response time and difficulty levels of the games in order to better suit the various patients with different abilities of upper extremity functions. b) To expand the sensor's sensing scope. c) To be

able to record movement data. d) To provide a Chinese version of the software interface of the games. e) To improve the ways to fix the controller on the user's hand. f) To fit the controllers size for different hand dimensions of the patients. g) To provide better correspondence between the game and real-life movements. h) To provide controllers for body control training, such as chest strap and belt. i) To simplify the controller's operation.

In order to make these devices more suitable to use in rehabilitation, a comprehensive follow-up design development based on these proposed guidelines would be necessary. The newly designed devices would be more suitable for rehabilitation therapy, and enhance the use safety and effectiveness for people of all ages and abilities.

ACKNOWLEDGMENT

This study is supported by the National Science Council of the Republic of China with grant No: NSC 100-2221-E-040-009-MY2.

REFERENCES

[1] L. L. Huang, C. F. Lee, and M. H. Chen, “The investigation of design improvement requirements for the upper extremity rehabilitation devices in Taiwan”. Proc. Asia Pacific Industrial Engineering & Management System Conference (APIEMS 2010), Dec. 2010.

[2] G. Saposnik, M. Levin, and stroke outcome research Canada working group, “Virtual Reality in Stroke Rehabilitation: A meta-analysis and implications for clinicians,” Stroke, vol.42, 2011, pp.1380-1386.

[3] L. Y. Joo, T. S. Yin, D. Xu, E. Thia, P. F. Chia, C. W. K. Kuah, and K. K. He, “A feasibility study using interactive commercial off the shelf computer gaming in upper limb rehabilitation in patients after stroke,” Journal of Rehabilitation Medicine, vol.42, no.5, 2010, pp.437-441.

[4] Y. J. Chang, S. F. Chen, and J. D. Huang, “A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities,” Research in Developmental Disabilities, vol.32, no.6, 2011, pp.2566-70.

[5] B. Lange, S. Flynn and A. Rizzo, “Initial usability assessment of off-the-shelf video game consoles for clinical game-based motor rehabilitation,” Physical Therapy Reviews, vol.14, no.5, 2009, pp.355-363.

[6] M. R. Mouawad, C. G. Doust, M. D. Max, and P. A. McNulty, “Wii -based movement therapy to promote improved upper extremity function post-stroke: a pilot study,” Journal of Rehabilitation Medicine, vol. 43, no.6, 2011, pp.527–533.

[7] G. Saposnik, R. Teasell, M. Mamdani, J. Hall, W. McIlroy, D. Cheung, K. E. Thorpe, L. G. Cohen, M. Bayley, and the Stroke Outcome Research Canada Working Group, “Effectiveness of virtual reality using Wii gaming technology in stroke rehabilitation: a pilot randomized clinical trail and proof of principle,” Stroke, vol.41, 2010, pp.1477-1484.

[8] G. Yavuzer, A. Senel, M. B. Atay, and H. J. Stam, "Playstation eyetoy games" improve upper extremity-related motor functioning in subacute stroke: a randomized controlled clinical trial,” vol.44, no.3, 2008, pp.237-44.

[9] J. Nielson, Usability Engineering, CA:Morgan Kaufman Publishers, 1993.

[10] P. Langhorne, F. Coupar, and A. Pollock, “Motor recovery after stroke: a systematic review,”Lancet Neurology, vol.8, no.8, 2009, pp.741-754.

[11] L. F. Lucca, “Virtual reality and motor rehabilitation of upper limb after stroke: a generation of process?,”Journal of Rehabilitation Medicine, vol.41, no.12, 2009, pp.1003-1100.

[12] Nintendo Wii, “Game Instruction Manuals and Wii System and Accessory - Instruction Manuals,” 2012, http://www.nintendo.com/consumer/manuals/index.jsp

[13] XaviX Port, “User's guide,” 2012, http://www.XaviX.com/products/manual/port_ug.pdf

[14] A. R. Fugl-Meyer, L. Jaasko, I. Leyman, S. Olsson, and S. Steglind, “The post-stroke hemiplegic patient. 1 . a method for evaluation of physical performance,” Scandinavian Journal of Rehabilitation Medicine, vol.7, no.1, 1975, pp.13-31.

[15] P. W. Duncan, M. Propst,and S. G. Nelson, “Reliability of the Fugl-Meyer assessment of sensorimotor of sensorimotor recovery following cerebrovascular accident,” Physical Therapy, vol.63, 1983, pp.1606-1610.

[16] V. Mathiowetz, G. Volland, N. Kashman, and K. Weber, “Adult norms for the Box and Block Test of manual dexterity,”The American Journal of Occupational Therapy, vol.39, no.6, 1985, pp.386-391.

[17] L. Brosseau and C. Wolfson, “The inter-rater reliability and construct validity of the Functional Independent Measure for multiple sclerosis subjects,” Clinical Rehabilitation, vol.8, 1994, pp.30-41.

[18] K. J. Ottenbacher, Y. Hsu, C. V. Granger, and R. C. Fiedler, “The reliability of the functional independent measure: a quantitative review,” Archives of Physical Medicine and Rehabilitation, vol.77, no.12, 1996, pp.1226-1232.

[19] J. F. Ravaud, M. Delcey, and A. Yelnik, “Construct validity of the functional independent measure (FIM): questioning the unidimensionality of the scale and the “value” of FIM scores, “ Scandinavian Journal of Rehabilitation Medicine, vol. 31, no.1, 1999, pp.31-41.

[20] C. V. Heck, I. E. Hendryson, and C. R. Rowe, “Joint motion: method of measuring and recording,”American Association of Orthopaedic Surgeons, 1965.

[21] C. Y. Lee, C. Y. Wu, I. N. Lien, M. H. Hsu, and K. H. Lin, “Effects of modified constraint-induced movement therapy on stroke patients,”Formosan Journal of Medicine, vol.10, 2006, pp.429-437.

[22] J. Cohen, “Statistical power analysis for the behavioral sciences,” Second Edition, Lawrence Erlbaum Associates, Hillsdale, 1988.

Evaluating the Interaction of Users with Low Vision in a Multimodal Environment

Clodis Boscarioli
Department of Computer Science
Western Paraná State University (UNIOESTE)
Cascavel, Paraná, Brazil
clodis.boscarioli@unioeste.br

Marcio Seiji Oyamada
Department of Computer Science
Western Paraná State University (UNIOESTE)
Cascavel, Paraná, Brazil
marcio.oyamada@unioeste.br

Jorge Bidarra
Department of Computer Science
Western Paraná State University (UNIOESTE)
Cascavel, Paraná, Brazil
jorge.bidarra@unioeste.br

Marcelo Fudo Rech
Department of Computer Science
Western Paraná State University (UNIOESTE)
Cascavel, Paraná, Brazil
marcelofudo@hotmail.com

***Abstract*—This paper presents the PlatMult environment, a multimodal platform that aims to provide accessibility in interactive kiosks to users with low vision and elderly ones. The PlatMult solution is composed of a screen magnifier (visual stimulus), a screen reader (auditory stimulus), and motor feedback (tactile feedback). The evaluation of the interaction with users with low vision is described, focusing on usability and accessibility aspects. This paper also discusses the potential of the platform for social and digital inclusivity. We conclude that the PlatMult, with its integrated features, helps users to access and use information systems in a suitable way.**

***Keywords*—PlatMult environment; low vision; usability; accessibility.**

I. INTRODUCTION

The advances made in Information and Communication Technologies (ICT) have resulted in benefits to society in many fields like industry, services, and even social inclusion.

ICT provides facilities to citizens, especially with regard to accessing, publishing, and sharing information. As an example, we can mention the automatic teller machines (ATMs) and interactive kiosks available in banks, bus terminals, airports, and libraries.

Despite all the progress that has been made, there are some people who cannot benefit from these technologies. Although many of the technologies are useful for people with disabilities, these people tend to have less access to these solutions. This work focuses particularly on those with visual impairments.

Among people with visual impairment a large contingent falls in the group with low vision. According to the classification of the World Health Organization (WHO), low vision refers to individuals who can only see an image at 20 feet (about 6m) where a person with normal vision would see it at 200 feet (about 60m). For Corn and Erin [5], low vision is a condition characterized by vision lag, where correction or improvement cannot be achieved by surgical means or solved with the use of conventional glasses.

Assistive Technologies (ATs) have emerged as an attempt to guarantee the same access rights to these people. AT, according Cook and Hussey [6], is a wide range of equipment, services, strategies, and practices designed and implemented to mitigate the problems faced by individuals with disabilities.

The implementation of hardware and software components, with increasing consideration given to the accessibility, usability, and adaptability features in their solutions, is now a reality. The challenge has been to seek simpler interfaces which can be immediately assimilated and handled correctly by users.

This work proposes a multisensory platform called PlatMult, implemented under the free platform philosophy, which is oriented towards but not restricted to people with low vision. Here we present some of the main results and discussions emerging from tests that we have conducted on our solution involving users with low vision. The methodologies adopted in the evaluation of PlatMult are based on the Nielsen's Heuristics [10] and usability principles. We therefore, seek to identify and correct the problems and gaps found in its interface. As a result of the interaction evaluation, changes are proposed to provide a better user experience.

This paper is organized as follows. Section II presents some related works. In the Section III the PlatMult platform is described. In Section IV, a brief description of the evaluation methods used is given. In Section V, the methodology used is described, and in Section VI the main results obtained up to now are discussed. The conclusions as well as future works are presented in Section VII.

II. RELATED WORKS

iBrowse interface software [12] was developed to help visually impaired people to access Internet. The software, adopted a similar designing strategy of the previously implemented LowBrowse software which acted as an extension add-on of Firefox browser. Its implementation, instead of following the traditional magnification technique, allows the low vision users to adjust parameters, such as font size, color and spacing, and then to read all the websites in their maximum reading efficiency. The tool provides also a screen reader using the traditional text-to-speech (TTS)

 ISBN: 978-1-61208-250-9

technology. In terms of implementation this solution uses the same Firefox plugin strategy we have adopted in our implementation but differently of our proposal changes the webpage content and do not furnish interaction motor feedback.

Sandhya and Devi [14] evaluate the level of accessibility in AJAX content and JAWS screen reader. AJAX stands for Asynchronous JavaScript and XML. AJAX allows feature-rich, dynamic web applications which use server-side processing without requiring the traditional submit or retrieve webpage methodology. If the screen reader is not prepared to this asynchronous updating, it will harm the user interaction. Moreover, the screen reader is limited to read images, advertisements banners and big data tables. Despite the authors list some recommendations to make screen reader accessible in webpages content, no user evaluation was done.

Tae et al. [8] propose an interactive kiosk for blind people with tactile interface (braille) and voice to identify images and characters on the screen area. Saito et al. [13] present flexible solution for blind and deaf people. Using a touch screen interface it can magnify images in different sizes and the both use voice messages and braille input. Tichy and Steinbrunner [15] have developed a tactile feedback for cursor control device with haptics.

Differently from the above cited works, this paper presents and evaluates the user interaction with the PlatMult platform, a tool developed specially to provide accessibility for low vision people. The main characteristic of this solution is the use of three different axes of interaction: visual, auditory and tactile.

III. THE PLATMULT PLATFORM

PlatMult [3] is an integrated solution of hardware and software developed to provide a multisensorial environment in interactive kiosks and ATMs, providing accessibility for users with low vision and elderly users. The solution (Figure 1) uses three areas of interaction: visual, audio, and tactile, which are integrated into a single solution, making it possible to provide users with better experiences.

One of this project's requirements is a low cost, to allow access to a great number of people. To assemble our prototype we used parts of computers and cell phones discarded by people and companies. The carcasses of slot machines were used as shown in Figure 2. It is important to highlight that slot machines are prohibited in Brazil, so the ones used in this project were donated by the Brazilian courts. The hardware platform uses a desktop computer with the following configuration: a Celeron 2 GHz processor with 1 GByte RAM and a 40 GByte hard drive. This configuration uses restricted hardware resources with outdated processors which can also be obtained through donations from organizations.

Additionally, the entire implementation was based on the free software philosophy. Ubuntu is the operating system used in this project; the windows manager is GNOME and the graphics server; see Xorg [16]. For accessibility features, the Assistive Technology Service Provider Interface (AT-SPI) library [1] is used. In this project, all the developed applications are free.

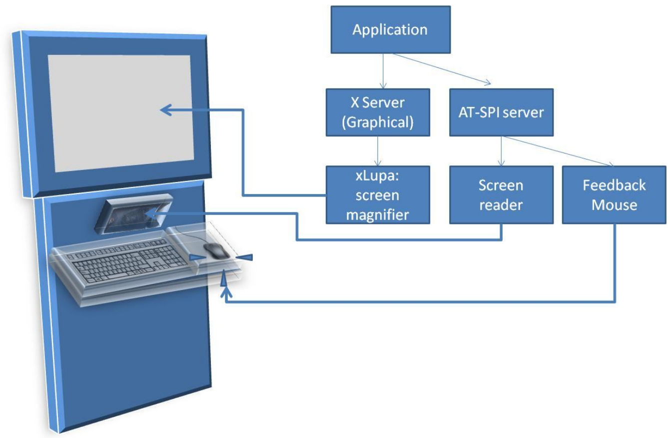


Figure 1. Overview of the PlatMult environment with the visual, auditory, and tactile components

Figure 2. Prototype of the PlatMult

The component responsible for providing the visual stimulus is the xLupa, an adaptative and free screen magnifier. The characteristics of the screen magnifier are full screen magnification and configuration of brightness, contrast, and the magnification factor. Additionally, the xLupa provides a mouse configuration, enabling the user to increase the pointer size or even change to a cross-pointer type.

 ISBN: 978-1-61208-250-9

Besides the visual features, there is also a screen reader to read texts as well as menu items, buttons, windows, and so on, based on the events reported by the accessibility Application Programming Interface (API) or AT-SPI. The screen reader was designed to warn the user about his or her actions, such as removal or insertion of characters in a text. The voice synthesis process is accomplished by eSpeak free-software; see [7].

The tactile stimulus is provided using a mouse with a motor feedback feature. The tactile feedback is activated when the mouse pointer is over menu items, links, and figures inside a webpage.

The feedback circuit in the mouse is constructed at low cost and aims at ease of implementation. To vibrate, the mouse uses the same principle as cell phones, which means that there is an engine with unbalanced weights on the axis. Such vibration occurs when the engine is activated and the weight causes the device to vibrate. Starting the engine requires 5 V and thus it is possible to use a parallel or USB port. So, the adapted mouse is connected to the PS/2 port (for normal operation) and to the parallel or USB port to receive vibration commands. Figure 3 shows a mouse that was built in this project. The vibration engine is glued to the mouse to allow tactile feedback.

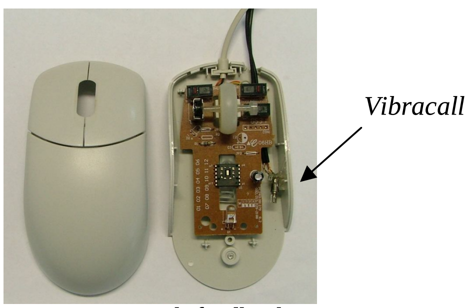


Figure 3. Mouse with feedback motor

The tactile server is implemented according to the same principle as the screen reader and uses the events generated by the AT-SPI.

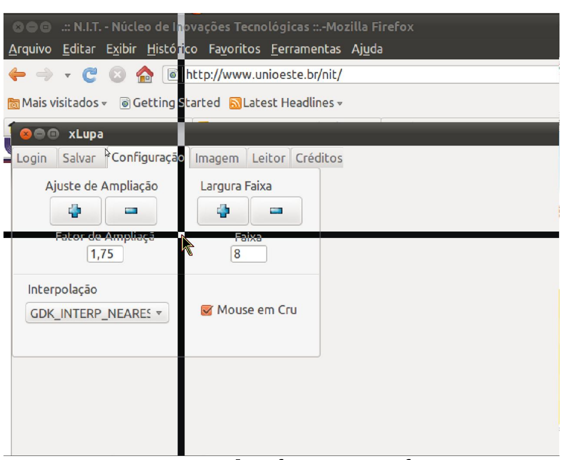


Figure 4. Example of xLupa configuration

Figure 4 illustrates one configuration screen for adjustments of the magnification factor and mouse size that permit user adaptation to environment.

IV. Evaluation Methods

The human–computer interface (HCI) evaluation is the process where the evaluator judges the quality of use of an HCI solution and aims to identify problems in the interface which may spoil the user's experience.

Due to several factors, such as the gathering bad or wrong requisites and implementation problems, the final result of a development process is not always a good quality product. The evaluation of these products makes it possible to release higher quality products.

During the software development process it is possible to identify two very distinct groups of actors involved. The first group comprises the people responsible for the final product concept, and the second, the people who will use the software. Because of this, an evaluation is performed from two different perspectives: the developers' and the users' perspectives, to guarantee that the product will be of higher quality.

From the users' perspective, the interaction with the system and how to achieve their goals will be evaluated. Aspects like usability, accessibility, and the interface's communicability issues need to be considered.

There are various methods of HCI evaluation, which can be classified as investigation, observation, and inspection described for Barbosa and Silva [2], and Rogers, Sharp and Preece [17]. The investigation methods normally allow the evaluator to find out the users' opinions and expectations of a system. On the other hand, the observation methods provide data about usage situations, allowing the identification of real problems. Finally, the inspection methods allow the evaluator to predict the consequences of design decisions.

The evaluations performed on PlatMult were an inspection evaluation using Nielsen heuristics and an observation evaluation using a usability test. As a first inspection evaluation and because the team expertise in Nielsen's Heuristics, we opted for this method, although other methods of evaluation are being explored to continue the evaluation process.

The use of Nielsen's set of ten heuristics is a method that was created to find usability problems during the design process; see [10] and [11]. The severity rating for each problem takes into account three factors: frequency, impact and persistency.

Given these three factors, the severity scale varies from a cosmetic problem to be solved if there is enough time in the project's schedule to a catastrophic problem which needs to be solved before the end of the project.

The usability test is an evaluation method where real users evaluate the software's interface according to usability factors like ease of learning, task memorization, user satisfaction, productivity, and error prevention.

Usability tests aim to evaluate the usability quality present in software, mainly by evaluating the user's performance when using the software [17].

 ISBN: 978-1-61208-250-9

With the usability test it is possible to evaluate the user's success rate when performing a task and the number of mistakes made, allowing those mistakes to be classified. Basically, according to Barbosa and Silva [2] and Rogers, Sharp and Preece [17], a usability test is composed of the following tasks:

(i) Preparation, which includes a creation of a guide containing the tasks which must be performed by the participants, recruiting a group of appropriate users, preparing the material for observation and recording, and finally executing a pilot test.

(ii) Data gathering, which aims to observe and register the performance and opinions of the participants during the controlled periods of use.

(iii) Interpretation and (iv) consolidation of results. This step includes gathering, counting, and summarizing the data collected from the participants and compiling a report of the results.

V. METHODOLOGY

PlatMult as a whole was inspected by three evaluators based on the Nielsen heuristics in an attempt to identify violations of these heuristics, which were documented according to the aspect, location, and degree of severity. For each violation, possible solutions detected for the problem are proposed.

Since the initial platform conception, experts such as ophthalmologists and teachers who work directly with students with low vision, as well as a group of low vision users participated in the process of requirements elicitation and interaction design. For the user tests presented here, two users were selected given their stereotypes. This small group was chosen to enable intensive and controlled evaluation, in order to proof the concept. All tests were followed by the experts and developers. In the future works, other usability tests will be proposed, involving other users.

Two users participated in the validation tests performed here, both of whom have low vision and significant hearing deficiency. They have beginner-level ability in the use of the Linux Operating System and the screen magnifier xLupa incorporated in PlatMult. A teacher specialized in special education also helped the testers and the appraisers to carry out the tests during the experiment.

Although, at first glance, the number of testers may seem low, the results obtained are expressive because the users have multiple deficiencies and are young. One is a teacher of students with visual deficiencies in the public education system and the other is a student enrolled in the special students class maintained by the public government. This student, besides being blind and deaf, has a severe debility of his right hand which obliges him to use his left hand despite being right-handed.

It is worth mentioning that although they are both users with some knowledge of computational environments, they have undergone training in the use of the xLupa magnifier delivered by other members of the project. In both cases, outside of the university or school where the project is being developed, their use of computers is infrequent, and this was their first contact with a screen magnifier. One of the users frequently uses a computer with a voice synthesis solution called DOSVOX, described for Borges in [4].

The guide which was followed in the experiment was elaborated so as to reflect everyday use and the proposed use of PlatMult. During the tests, the users completed a series of tasks including an Internet search scenario. For each task, we were able to evaluate the interaction, identifying problems encountered in utilizing the platform. The results include the total number of errors and the time needed to complete each given task.

The test guide was divided into two parts, the first with the xLupa activated and the second with all three features (visual, tactile, and auditory) available on PlatMult activated.

A. Usability Test – Scenario 1

In the first part, the following tasks were given to the users:

a) Activating xLupa, preferably without the automated initialization feature activated. The aim was to test whether the user was able to identify one of the two possible ways to initialize the screen magnifier.

b) Setting up his or her profile on xLupa, selecting the magnification factor and the color pattern which were more suited to him or her. This test also involved configuring the mouse pointer and applying the antialiasing algorithm to the text and images.

The objective of this task was to verify whether the user was able to manipulate the menus for changing the magnification factor and color. The task provided answers to the following questions: Is the user capable of identifying which factor and/or patterns have been selected? Is he or she able to differentiate the selected configuration from the other available ones?

These two initial tasks, (a) and (b), were also used to verify whether the basic configuration applied was adequate for a good experience of the utilization of xLupa or whether it was necessary to make some adjustments to the settings, which had to be set up every time until they were considered satisfactory. The objective was to verify whether the users had sufficient knowledge to perform such adjustments.

c) Initializing the browser Firefox, accessing Google's webpage, and searching for information specified by the evaluation team.

This task was done to identify the degree of ease/difficulty of use, and the users performed the following tasks:

d) Finding the Firefox URL address bar;
e) Typing in the desired URL address;
f) Finding Google's search field;
g) Finding the link indicated by the evaluation team;
h) Finding the solicited information on the webpage.

B. Usability Test –Scenario 2

The stages of the second usability test were very similar to those of the first; however, the steps related to the xLupa's configuration were not repeated, and the users had to perform the following tasks:

a) Activating the screen reader on xLupa's configuration window.

This task aimed to determine whether the users had enough knowledge of how to activate the screen reader, and if so, to determine the degree of ease/difficulty with which they performed the task. If they were not able to do it, the study aimed to find out the reason why.

b) Initializing Firefox, and, with the plugin and screen reader activated, accessing Google and searching for different requested information so that the process would not be repetitive. These actions were performed with the objective of observing how these features influence the user's experience.

One of the first actions performed by the users was based on the initial configuration of the profile of each user and comprised adjusting the tool's functionality, more specifically the magnifier, and their visual needs.

The participants, aided by the teacher who watched them, answered a questionnaire post-test inquiring about the utility of the environment, difficulties with the interactions, their motives, and alternative ways to make the platform easier to use. Questions were also asked about the prototype's ergonomics.

VI. RESULTS AND DISCUSSION

This section presents the results gathered after the application of the inspection and observation methods.

A. *Nielsen heuristics evaluation*

In this section, only the violations of the Nielsen heuristics will be reported.

User control and freedom: A problem occurs at the visual acuity selection screen and color scheme selection. If the user selects an inappropriate magnification factor and/or color scheme, he or she will have some difficulty solving the problem because the only way to do so is to close the magnifier, restart it, and then redo the configuration process. This was classified as a problem because this process tends to be hard for users with low vision to perform without help.

The results showed, however, that this difficulty may be considered cosmetic, given the fact that this functionality is accessed during the profile's initial configuration, and with the automated loading of the last saved profile, it becomes a minor problem. With regard to the platform itself, however, the most appropriate solution would be to add a button giving the user the option to restore the default settings without closing the software.

Another violation of this heuristic was found under the "Save" tab with regard to the "Remove" button, which executes the operation as soon as the button is clicked without giving the user the option to cancel the action. As a solution to this violation, the addition of an alert window asking the user to confirm the action is suggested.

Consistency and standards: One problem was found at the configuration screen. On the "Save" tab, the options "Save" and "Update" perform the same operation, and because of this, the user may get confused as both operations store current user information. The problem was classified as small, and the solution to this problem would be to automatically update the user configurations, rendering the "Update" button useless; therefore, the "Update" button would be removed and only the "Save" button would be left, which would only be used to create a new user profile.

Match between the system and the real world: One problem in this category was detected at the configuration screen. Under the "Configurations" tab, it was difficult for the user to understand the meaning and effect of the feature "Strip width". The problem was classified as small, considering that the user does not use this feature often. A possible solution would be to name the feature differently, for example as "Cross-pointer size".

Another violation of this heuristic can be found at the configuration screen, under the "Configurations" tab, with regard to the selection of the type of Interpolation. The terms which identify the types are technical and are related to the implemented framework. The problem was classified as small. A simple change of names and descriptions to less technical terms would be sufficient to fix this problem. Another possible solution would be to move these options to a new tab called "Advanced Configuration", since it is a feature that is accessed by individuals who aid users with low vision.

Another problem at the configuration screen was found in the "Image" tab. A violation was found in that the term "Image" does not inform the user about what types of configurations this feature handles. This problem was classified as small, and could be easily fixed by changing the tab name to "Image settings'.

Another problem was seen at the configuration screen under the "Image" tab, where the labels on the checkboxes "Change theme" and "Gray" were not clear to the users. The problem was classified as small and could be fixed by a simple change of the names to "High contrast" and "Shades of gray".

Another violation of this heuristic was found at the configuration screen, under the "Reader" tab, with regard to the buttons "Play" and "Stop", which are both written in English. However as the program will be used by Portuguese-speaking users, the labels should be translated into Portuguese.

At the configuration screen, under the "Reader" tab, the label "Speed" is too vague and does not represent the function properly. The solution would be to change the label to "Reading speed" or "Reader's speed".

This was the first inspection performed of the interface of xLupa, and it could be very helpful to allow the developer team to make improvements for users with low vision. Furthermore, it is through the xLupa's interface that the user interacts with the system.

B. *Usability test results*

To complement the evaluation, the usability test was performed to verify the performance and level of satisfaction of the two real users with low vision when they interacted with the PlatMult platform to execute everyday tasks as mentioned above.

During the tests, the evaluators reported that the mouse pointer was too small before the initialization of xLupa, which would cause difficulties for users with low vision. In

 ISBN: 978-1-61208-250-9

addition, the speed of movement of the pointer was another difficulty.

There were moments at which both testers got lost in the middle of the information shown on the screen. The main reason for this was the vertical rolling, an essential feature in an amplified screen, where not all of the information will fit on the screen. In situations like this, it is necessary to carry out user training.

The evaluation team had difficulty understanding what the screen reader was saying in some cases, especially when browsing the Internet.

Another problem was the operating system update notifications, which overlapped the information that the user might have wanted to find.

The users reported that the motor feedback mouse was very useful, especially for browsing the Internet, because it helped in identifying useful information on screen.

As for the screen reader, the users reported that it was very useful, mainly when typing text, and they did not report any problems with it. Finally, it is worth mentioning that this analysis works being done with these students prioritizes not only visual stimuli but also pedagogic tasks and cognitive aspects.

The students' interaction with the group was positive, because they were submitted to basic informatics concepts, memorization exercises, and reading and writing texts. It was observed that these exercises contributed to developing the visual part and main focus of our work.

VII. CONCLUSION AND FUTURE WORK

This paper presented the process and results of a usability evaluation of the PlatMult environment by inspection and observation, which is important in the current state of the project, where improvements to the interaction and interface will be made.

In the tests, both users with low vision mentioned the mouse pointer speed as a problem. One suggested way of fixing this problem would be to integrate an option to configure the mouse speed on xLupa's interface.

The screen reader proved useful and efficient. Despite this, it is recommended that users with low vision should not use screen readers, to avoiding discouraging them from training their vision.

With this experiment, it is possible to conclude that the PlatMult, with its integrated features, helps users to access and use information systems in a suitable way.

Although these tests were performed to evaluate the features of the tool's accessibility, it is worth mentioning that investment in this platform is justified by its different areas of applicability. The present solution can be useful in ATM terminals of banks and also in sectors of governmental services like public libraries and schools.

In future works, some observed violations of the Nielsen heuristics will be fixed in the interface. The interactive kiosk is then expected to have a high degree of accessibility for the final user. During the usability tests, one of the difficulties identified was the mouse pointer size and speed before the activation of the screen magnifier. One solution, using shortcut keys to activate and configure some parameters of the screen magnifier, is being developed by the team. Additionally, the magnifier can be configured to be initialized together with the operating system startup, removing the difficulties of running xLupa.

Others inspection methods such as Cognitive Walkthrough described in [9], and ISO 9241-210 will be better investigated and applied.

Another point which was identified during the test was the notification interface of the operating system. In this case study, the operating system was run using its default configurations, which included alerts about package updates. With the aim of standardizing and easing installation on new machines, a new repository will be set up containing the packages and an operating system installer which includes all the accessibility features of PlatMult. The operating system will be configured so that alerts are not shown while the system is being used.

With regard to the usability tests, we intend to test new scenarios, mainly including the use of Internet services, such as post office services and government health care services. Additionally, scenarios involving the use of educational software will be developed and applied in usability tests.

We also intend to increase the number of users participating in prototype testing in order to have more coverage from the perspective of users with different characteristics.

With regard to the motor feedback, the initial evaluation, which primarily concerned content in the Internet browser, was positive. A second version of the mouse containing two engines positioned on the right and left sides is under construction. With these two engines, we intend to create vibration patterns which can indicate the importance of the information at which the mouse is pointing and to try to create a spatial stimulus, aiming at a fast localization of the pointer on the screen.

VIII. ACKNOWLEDGMENTS

We are grateful to the Araucaria Foundation and CNPq for funding the project and we would like to thank all the people involved in this research.

REFERENCES

[1] AT-SPI-Gnome, Available at http://acessibility.gnome.org. [retrieved: December, 2012].

[2] S. D. J. Barbosa and B. S. Silva, Interação Humano-Computador. Série SBC, Campus-Elsevier, 2010.

[3] J. Bidarra and M. S. Oyamada, "Development of an interactive kiosk with screen amplifier targeting low vision and old-aged people", Proc. AIRTECH – Accessibility, Inclusion and Rehabilitation using Information Technologies, Havana, Cuba, December. 2011, pp. 13-14.

[4] J. A. Borges, Dosvox – um novo acesso dos cegos à cultura e ao trabalho. Revista Benjamin Constant, Rio de Janeiro, v. 3, May. 1996, pp. 24-29.

[5] A. L. Corn and J. N. Erin, Foundations of Low Vision: Clinical and Foundation Perspectives. 2nd Ed. AFB Press: American Foundation of the Blind, 2010.

[6] A. M. Cook and S. Hussey, Assistive Technologies: Principles and Practices. St. Louis, MO: Mosby-Year Book, Inc., 1995.

[7] eSpeak. Available at http://espeak.sourceforge.net/. [retrieved: December, 2012].

[8] J. B. Tae, L. H. Ju, L. U. Taek, P. C. Yong and S. M. Gyu, Information terminal for blind person based on tactile and voice interface. Patent Number: KR 20010017729. Mar. 2001.

[9] J. Bradford, C. Wharton, J. Jeffries and M. Franzke, “Applying Cognitive Walkthroughs to more Complex User Interfaces: Experiences, Issues and Recommendations”. CHI ’92 - Proceedings of the SIGCHI Conference on Human Factors in Computing. Systems, Monterey, California, USA. 1992. pp. 381–388. doi: 10.1145/142750.142864

[10] J. Nielsen, Usability Engineering. San Francisco, USA: Morgan Kaufmann, 1993.

[11] J. Nielsen and R. L. Mack, Usability Inspection Methods. New York, USA: John Wiley & Sons, 1994.

[12] L. Jianwei, X. Guobei, X. and Guotong, “iBrowse: Software for low vision to access Internet”, Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on, vol.5, Yantai, China. 16-18 October. 2010. pp. 2062-2066. doi: 10.1109/BMEI.2010.5639862

[13] M. Saito. Barrier-free terminal. Patent Number: JP11184368. Dec. 1999.

[14] S. Sandhya and K.A.S, Devi, “Accessibility evaluation of websites using screen reader”, Next Generation Web Services Practices (NWeSP), 2011 7th International Conference on, Salamanca, Spain. 19-21 October. 2011. pp. 338-341. Doi: 10.1109/NWeSP.2011.6088143.

[15] T. H. Tichy and S. L. Steinbrunner. Tactile feedback for cursor control device. Patent Number: US6967643. Nov. 2005.

[16] Xorg. Available at http://linux.die.net/man/1/xorg. [retrieved: December, 2012].

[17] Y. Rogers, H. Sharp, and J. Preece. Interaction Design – Beyond Human-Computer Interaction. 3rd Ed. John Wiley Professional, 2011.

Bimanual Performance in Unpredictable Virtual Environments

A Lifespan Study

Andrea H Mason, Drew N Rutherford, Patrick J Grabowski
Department of Kinesiology
University of Wisconsin - Madison
Madison, USA
amason@education.wisc.edu, anrutherford@wisc.edu, pgrabowski@wisc.edu

***Abstract*—Interaction and interface design for the young and the elderly has become an important research topic. The purpose of the research described here is to characterize motor performance in virtual environments across the lifespan. Participants between the ages of 7 and 90 years simultaneously reached to pick up two objects with their right and left hands in a desktop virtual environment. On random trials, objects were unexpectedly moved to new locations. Results indicated that older adults used different movement strategies in the virtual environment when compared to results from natural environment experiments. Further, children and older adults responded to perturbation conditions with different movement time and hand coupling strategies than young and middle-aged adults. These results suggest that age and task-specific design is necessary to ensure general access and optimal performance in virtual environments.**

Keywords- virtual environment; aging; motor control; bimanual reach to grasp

I. Introduction

A. HCI and Age

With the expansion of the role of computers in schools, the workplace, and homes, the population of users who make regular use of computing technology has grown exponentially. Unfortunately, Human-Computer Interaction (HCI) research has not reflected this demographic reality. Results of the 2010 US Census show that 17.5% of the US population is between the ages of 5 and 18 and a further 40% of the population is above the age of 45 [1]. It has also been reported that Europe is experiencing an aging population, with projections of 35% of the population being above the age of 65 by 2025 [2]. Still most HCI research is focused on younger people, often university or college students [3]. Rather than representing the true population of computer users, most experimental HCI research is biased heavily towards the cognitive and motor abilities of young adults. Where age-specific research has been conducted, the majority relates to the design of standard computer interface systems for various age groups. In particular, research has focused on ways to improve cognitive performance through specific training or tutorial methods [e.g. 4, 5], or on the age-appropriate design of input devices [e.g. 6, 7, 8].

While a modest corpus of knowledge is available for the design of standard computer interface systems for a variety of age groups, much less is known about how age influences performance within immersive three-dimensional virtual environments (VEs) [9, 10, 11]. Immersive VEs are becoming more prominent as the costs of the relevant tracking and display technologies decrease. VEs are commonly used in design and prototyping, data visualization, medical training, architecture, and entertainment. Further, recent research has focused on the utility of VEs for rehabilitation of motor impairments such as stroke in the elderly and attention deficit hyperactivity disorder (ADHD), developmental coordination disorder and cerebral palsy in the young [12, 13]. However, because there is a paucity of information on how healthy children and older adults interact in VEs, it is likely that the success of these systems will struggle. Specifically, it is nearly impossible to extrapolate design characteristics from healthy young adults to special-needs children and older adults. Results of the few studies conducted on performance across age-groups within virtual environments indicate relevant disparities in reactions to environmental immersion, usage of various input devices, size estimation ability, and navigational skills [9, 10, 11]. According to Allen et al. [9], “these results highlight the importance of considering age differences when designing for the population at large.”

The purpose of the research described here is to characterize motor performance in virtual environments across the lifespan. To do this we asked participants ranging in age from 7 to 90 years to perform a foundational skill (bimanual reach to grasp) within a table-top virtual environment. In the following sections, we describe the importance of the skill we chose to study.

B. Bimanual Reach to Grasp Skills

The performance of many everyday activities requires the completion of asymmetric but coordinated movements with our two hands. For example, touch typing, tying our shoelaces, and even reaching for a mug with one hand and a coffee pot with the other require the performance of two separate but coordinated movements. Many asymmetric bimanual tasks such as the ones described above can be performed quite effortlessly in natural environments. This seamless control is possible because we use feedforward sensory information (vision and proprioception) to pre-plan our movements and feedback sensory information for on-line corrections during movement execution.

Recently, bimanual tasks have been targeted as important skills to (re)train in rehabilitation protocols employing natural environments and virtual reality [14]. In rehabilitation training after stroke, these types of tasks are important for functional recovery because they require the areas of the brain most commonly afflicted by stroke to work with areas usually left undamaged, thereby maximizing the potential for positive neuroplastic changes [15].

While the study of bimanual movements has received some attention in natural environments, very little is known regarding the performance of these types of movements in virtual environments [16]. Further, no studies have looked at how the control of bimanual skills changes as a result of age in VEs. In order to successfully implement rehabilitation and training protocols that make use of these types of tasks it is imperative that we first obtain a baseline understanding of how neurologically "normal" people across the lifespan perform bimanual skills in VEs and how they use sensory information for the performance of these skills.

In natural environments, results from bimanual movement studies have indicated that when the two limbs are used to accomplish both symmetric and asymmetric task goals, coupling between the limbs for certain parameters occurs in the temporal domain [17, 18]. In particular, movement onset, duration, and end times tend to be similar for the two hands when subjects aim toward or reach to grasp targets of different sizes or at different locations [17, 18]. However, timing differences between the hands have been shown, and results indicate that these differences are associated with insufficient visual feedback for movement control [19]. In the current study we investigated whether the same patterns of results are seen in virtual environments and whether these patterns change with age. We employed a target perturbation to specifically investigate how sensory (visual) information is used on-line by participants of various ages to modify their movements. These paradigms are discussed in more detail in the following section.

C. *Unpredictable Environments: Perturbation Paradigms*

An experimental paradigm that has been successfully used to investigate the role of on-line visual information for the performance of goal directed tasks uses target perturbation to study adjustments to ongoing movements. The use of this type of paradigm allows us to discern how long it takes the nervous system to adapt to an unexpected visual change as well as the efficiency of the adaptation.

In a target perturbation paradigm, the participant is unexpectedly presented with the requirement to alter their original movement plan either prior to or after movement onset. An example of a typical perturbation paradigm is as follows. A visual stimulus is presented to the participant prior to movement initiation and the participant generates a movement plan appropriate to the acquisition of the target at this initial location. Shortly prior to or after movement onset the stimulus is suddenly replaced by a second stimulus presented at an alternative location. The participant is thus required to reorganize their movement to successfully grasp the target at its new position. Results of studies using perturbation paradigms in both natural [20] and virtual environments [16] have indicated increased movement times to displaced targets and double velocity peaks in kinematic recordings.

Studying the performance of bimanual perturbation tasks in a VE can provide us with important information about how participants make use of visual information during the execution of a skill. This is particularly important given that the use of sensory information changes across the lifespan [21,22] and all the visual information presented to users of VEs must be synthetically created. By comparing results in the VE to studies performed in the "real" world we can determine whether performance is similar within these two environments.

II. METHOD

A. *Participants*

Fifty-one participants were divided into four age categories: Children (7-12 years, n=13), Young adults (18-30 years, n=12), Middle age adults (40-50 years, n=12) and Older adults (60+ years, n=12). Due to problems with data collection final data analysis was conducted on 12 participants in the "Children" group and 11 participants in the "Older adult" group. Decades of motor control research has indicated that a sample size of 10-12 participants provides sufficient statistical power in this type of reach to grasp study. All participants were self-reported right-handers and had normal or corrected-to-normal vision. All participants provided informed consent before taking part in the experiment. The protocol was approved by the University of Wisconsin-Madison Social and Behavioral Science Institutional Review Board.

B. *Experimental Apparatus*

This experiment was conducted in the Wisconsin Virtual Environment (WiscVE) at the University of Wisconsin-Madison. In this environment, subjects see three-dimensional graphical representations of target objects but interact with physical objects. As shown in Fig. 1, graphic images of two target cubes were displayed on a downward facing computer monitor. A half-silvered mirror was placed parallel to the computer screen, midway between the screen and the table surface. The graphic image of the cubes was reflected in the mirror and appeared to the participant to be located in the workspace on the table surface. Three light

Figure 1: Experimental apparatus

emitting diodes (LEDs) were positioned on the top surface of two wooden target cubes (38 mm). A VisualEyez 3000 motion capture system (Phoenix Technologies, Inc., Burnaby) tracked the three-dimensional position of the LEDs on the physical target cubes. This data was used with an 8–10 ms lag (which was not discernible to subjects), to generate the superimposed graphical representations of the cubes. A shield was placed below the mirror to prevent subjects from seeing the real environment or their hands as they performed the reach-to-grasp task.

Participants wore CrystalEYES™ goggles to obtain a stereoscopic view of the graphic images being projected onto the mirror. Three LEDs were fixed to the goggles and were used to provide the subject with a head-coupled view of the virtual environment on the work surface. Thus, when the subject moved his/her head, the displayed scene was adjusted appropriately for the magnitude and direction of head movement. LEDs were also positioned on the subject's right and left thumbs, index fingers and wrists. Data from all LEDs was collected at a sampling rate of 120 Hz and was stored for data analysis purposes.

C. Design and Procedure

Each trial began with the illumination of two blue circular start positions (radius 5 mm) located 12.5 cm to the left and right of the participants' midline. The participants moved their hands from the periphery of the workspace to place their index fingers and thumbs over the start positions, which were haptically indicated by small metal hex nuts. When the participants' hands were correctly positioned, the start positions turned yellow. Once both of the participants' hands remained stationary at the start positions for 1 s, the two graphic target cubes appeared at a location 20 cm from the start position. The task was to reach forward with the right and left hands to grasp and lift the two target cubes. Grasps were made with a precision grip and participants were asked to move at a comfortable pace once the target cubes appeared.

Participants experienced trials in four experimental conditions. In the control condition both targets remained at their initial location throughout the trial (left target no jump/right target no jump; NN). In the three perturbation conditions one or both targets were displaced 9 cm toward the participant at movement onset (defined as a displacement of 5 mm of the thumb LED). The perturbation conditions consisted of: 1) left target jump/right target no jump (JN), 2) left target no jump/right target jump (NJ), 3) left target jump, right target jump (JJ).

Participants performed a total of 100 trials. The first 10 trials were always control trials (NN). This allowed participants to become comfortable with the task and also gave us the opportunity to analyze a set of "control" trials where participants had no expectation of a perturbation. The remaining 60 control and 30 perturbation trials, 10 in each condition, were presented in a random order.

D. Data Analysis

Position data from the block LED as well as LEDs on the wrists of both hands were analyzed for specific temporal kinematic measures. Start of movement was defined as the point where wrist velocity increased above a threshold of 5 mm/s and continued increasing to a peak. End of movement was defined as the point where block lift velocity increased above 5 mm/s and continued increasing to a peak. Based on these two temporal measures we calculated Movement Time (MT) for both hands. We also quantified temporal coupling of the two hands by determining whether the hands started and ended movement at similar times. To do this we calculated the Absolute Start Offset (ASO: Start Left Hand – Start Right Hand) and Absolute End Offset (AEO: End Left Hand – End Right Hand).

Data were statistically analyzed in two ways. First, to quantify control performance in the first 10 trials, we conducted a 4 Group (Children, Young Adult, Middle Adult, Older Adult) X 2 Hand (left, right) repeated measures ANOVA on MT. To quantify bimanual coupling during the control trials a 4 Group (Children, Young Adult, Middle Adult, Older Adult) repeated measures ANOVA was performed on ASO and AEO. To quantify performance during the perturbation trials we conducted separate 4 Group (Children, Young Adult, Middle Adult, Older Adult) X 4 Condition (JJ, NJ, JN, NN) repeated measures ANOVAs for each hand and dependent measure. Post-Hoc analysis on significant main effects was done using the Fisher LSD method. When significant interactions occurred, these were further explored using simple main effects with Condition as the factor. An a priori alpha level was set at $p < 0.05$.

 ISBN: 978-1-61208-250-9

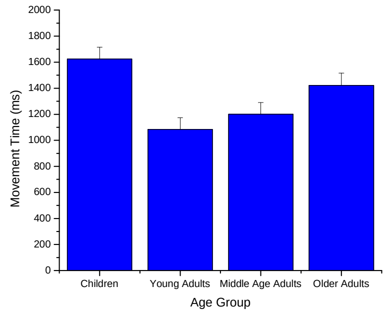


Figure 2. Main effect of Group on movement time in the control condition.

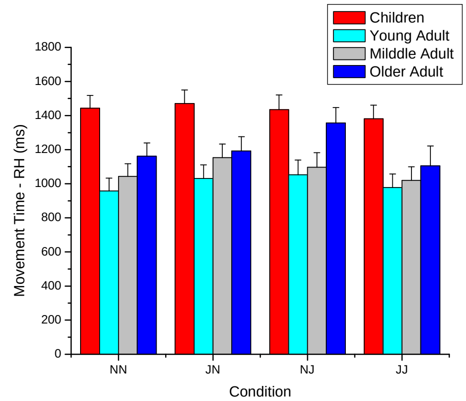


Figure 4. Group X Condition interaction for MT of the right hand.

III. RESULTS

A. *Initial Performance: Control Trials*

The control trials allow us to determine how bimanual performance changes as a function of age within virtual environments and whether patterns of performance in VEs replicate those seen in natural environments.

A main effect of Group was found for movement time ($F_{3,43}$ = 7.053, p=0.001). Results indicated that the fastest movement times were found in the young and middle aged adults. Children were significantly slower than the young and middle aged adults, whereas older adults were only significantly slower than the young adults (Fig. 2).

When looking at coupling between the left and right hands, main effects of Group were found for ASO ($F_{3,43}$ = 14.03, p<0.001) and AEO ($F_{3,43}$ = 4.74, p=0.006). The post-hoc LSD indicated that children had significantly larger offsets at both the start (Fig. 3A) and end (Fig. 3B) of movement than any of the other age groups.

B. *Perturbation Performance*

The perturbation trials allowed us to investigate whether differences in the use of on-line visual feedback occur across age groups and for different perturbation conditions.

An interaction between Condition and Group ($F_{9,129}$ = 2.934, p=0.003) was found for MT of the right hand. Children were significantly slower than all other groups in the NN, JN and JJ conditions (Fig. 4). However, they did have similar MTs to the older adults in the NJ condition. The young and middle adults had similar MTs across all conditions. Finally, the older adults were significantly slower than the young adults in the NN and NJ conditions only.

For MT of the left hand, main effects of group ($F_{3,43}$ = 6.04, p=0.002) and condition ($F_{3,129}$ =10.6, p<0.001) were found. The group main effect indicated that the children were significantly slower than the young and middle adults. No other significant differences were found (Fig. 5A). For the main effect of condition, results indicated that MTs for the left hand were significantly faster in the NN and JJ conditions than in the JN and NJ conditions (Fig. 5B).

When looking at coupling between the two hands during perturbation trials, a main effect of group ($F_{3,43}$ = 15.9, p<0.001) indicated that children had significantly larger offsets at movement initiation than any other age group (Fig. 6).

For the end of movement, a Group X Condition interaction ($F_{9,129}$ = 2.232, p=0.024) indicated that children had significantly larger offsets than all other groups in the NN condition (Fig. 7). The older adults had longer offsets than the young adults in the NJ condition. All groups had statistically similar offsets in the JN and JJ conditions.

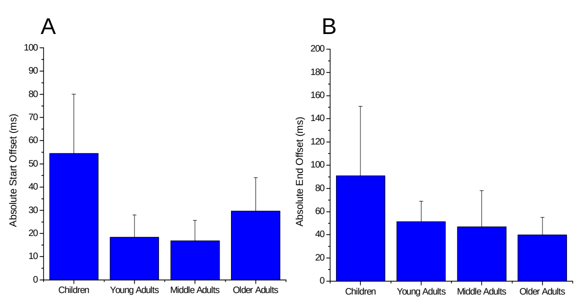


Figure 3. Main effect of Group on ASO and AEO

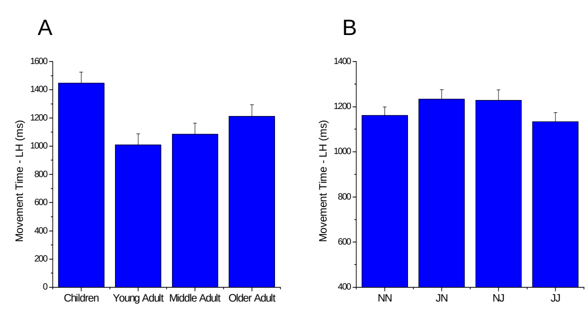


Figure 5. Main effects of Group and Condition on MT of the left hand

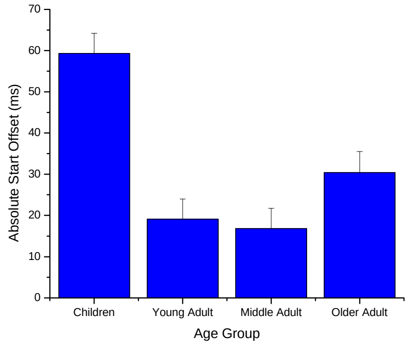


Figure 6. Main effect of Group on ASO

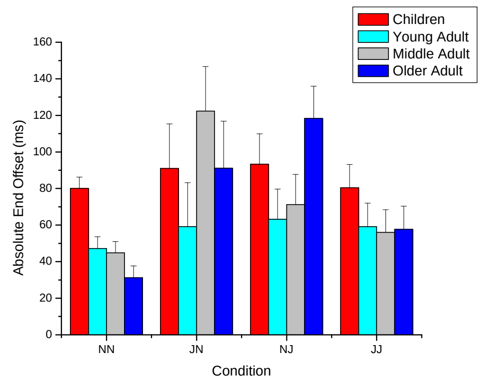


Figure 7. Main effect of Group on AEO

IV. Conclusion and Future Work

A. Performance of Bimanual Movements in VEs across the lifespan: Control and Perturbation Conditions

Each participant began the experiment by performing a block of simple bimanual trials without perturbation. These trials allowed us to determine whether age-specific patterns of bimanual performance in VEs are similar to the patterns seen in the natural environment. When considering overall MT, research in natural environments has indicated that children and the elderly typically complete both simple and complex tasks more slowly than young adults [23, 24]. A similar pattern of results was found in the current study, indicating some similarities between VEs and natural environments. With respect to bimanual coupling in natural environments, prior studies have indicated that both young children and older adults exhibit greater offsets at movement initiation and movement completion than young adults [24, 25]. These results were replicated for the children; however, the older adults used similar movement offset patterns as the young and middle adults. This difference in movement coupling for the elderly subjects suggests that they use different control strategies in natural compared to virtual environments. Timing differences between the hands in bimanual tasks have been associated with the requirement to shift visual attention between the targets to obtain sufficient feedback [19]. In older adults, slowing of visual sensory processing due to aging should result in even greater timing differences between the hands [22]. The smaller offsets seen in the current study suggests that the elderly subjects may have been relying on a predominantly feedforward strategy to complete the task instead of the typical feedback-based strategy that is seen in the natural environment. In a previous study investigating age differences on a simple reach-to-grasp task in a VE, we also found that older adults relied more heavily on a feedforward-based strategy [26]. The current findings add support to the notion that older adults may not rely on similar movement planning and execution strategies when performing tasks in VEs.

The perturbation conditions allowed us to investigate age differences in the visual control of movement in VEs. Overall, MT and offset results indicated similar movement performance between the ages of 18 and 50 years. These results suggest that design principles extracted from studies done on young adults may be applicable to middle-aged adults as well. In contrast, children and older adults exhibited distinct performance differences as a function of perturbation condition. While their performance was similar to the young and middle age groups for certain parameters and on certain conditions, the youngest and oldest age groups were slower and their movements were less coupled in other conditions. Overall, these results suggest that task conditions and age are critical factors when considering the design and functionality of VEs. Children and older adults do not perform or make use of sensory information in a similar fashion to young and middle-aged adults. Further, results are clearly task specific. This suggests that it is dangerous for designers to extrapolate performance in one task to other tasks. Instead, our results suggest that age-related performance must be investigated on a task by task basis for the generation of design principles.

B. Implications for the Design of Training and Rehabilitation VEs

Virtual environments have recently been touted as promising tools for training and rehabilitation [12, 13, 14]. However, the capacity for these environments to provide optimal benefits hinges on the learner's ability to transfer gains made in the VE to improvements in performance in the real world. It has long been known in the human motor learning literature that successful transfer occurs when similarities in movement strategies between the practice and performance environment are greatest [27]. In the current study we found that children, young, and middle-aged adults used similar bimanual strategies in the control condition to those reported for natural environments. In contrast the strategies used by the older adults in the VE were different than those reported in natural environments. It is important to note that visual feedback in this study was impoverished and relatively crude (i.e. no hand representation, simple table surface and object representation, low luminance contrast levels). These results suggest that when designing environments for older adults, it may be necessary to design tasks and environmental feedback conditions that better mimic the richness of the visual feedback conditions available in the real world. We are planning future studies to test this hypothesis.

ACKNOWLEDGEMENT

We would like to thank Brandon Bernardin, Stephanie Ehle, Nick Yeutter and Andrew Minkley for their assistance with data collection and analysis. This work was funded by the National Science Foundation, grant number 0916119.

REFERENCES

[1] L.M. Howden and J.A. Meyer, “Age and Sex Composition 2010: Census Brief,” Retrieved from URL http://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf.

[2] European Commision, “Eurostat Population Projections,” Retrieved from URL http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/main_tables.

[3] A. Dickinson, J. Arnott and S. Prior, “Methods for human-computer interaction research with older people,” Behav Inform Technol, Vol. 26, No.4, July-August 2007, pp.343-352, DOI: 10.1080/01449290601176948.

[4] K. Günther, P. Schäfer, B. J. Holzner, G. W. Kemmler, “Long-term improvements in cognitive performance through computer-assisted cognitive training: A pilot study in a residential home for older people” Aging Ment Health, Vol. 7, No. 3, May 2003, pp. 200-206.

[5] D. Hawthorn, “Interface design and engagement with older people,” Behav Inform Technol, Vol. 26, No. 4, 2007, pp. 333-41.

[6] M.W. Smith, J. Sharit, S.J. Czaja, “Aging, motor control, and the performance of computer mouse tasks,” Hum Factors, Vol. 41, No. 3, Sep. 1999, pp. 389-396.

[7] B. Laursen, B.R. Jensen, A. Ratkevicius, “Performance and muscle activity during computer mouse tasks in young and elderly adults,” Eur J Appl Physiol, Vol. 84, No. 4, Apr. 2001, pp. 329-336.

[8] A. Worden, N. Walker, K. Bharat, S. Hudson, “Making computers easier for older adults to use: area cursors and sticky icons,” Proc of the SIGCHI conference on Human factors in computing systems (CHI ’97), ACM Press, Apr. 1997, pg. 266-271.

[9] R.C. Allen, M.J. Singer, D.P. McDonald, J.E. Cotton, “Age differences in a virtual reality entertainment environment: A field study,” Proc. 44th Annual Meeting of the Human Factors and Ergonomics Society (IEA/HFES 2000), Published by the Human Factors and Ergonomics Society, July 2000, pp. 542-545.

[10] F.A. McCreary, R.C. Williges, “Effects of Age and Field-of-View on Spatial Learning in an Immersive Virtual Environment” Proc 42nd Annual Meeting of the Human Factors and Ergonomics Society (HFES 1998), Published by the Human Factors and Ergonomics Society, Oct. 1998, pp. 1491-1495.

[11] D.P. McDonald, D.A. Vincenzi, R.V. Muldoon, R.R. Tyler, A.S. Barlow, and J.A. Smither, “Examination of age-related performance differences on a locomotion task in a virtual environment (abstract),” Proc. 42nd Annual Meeting of the Human Factors and Ergonomics Society (HFES 1998), Published by the Human Factors and Ergnomics Society, 1998.

[12] M.K. Holden, MK, “Virtual environments for motor rehabilitation: review” Cyberpsychol Behav, Vol. 8, June 2005, pp. 212-219.

[13] M. Wang, D. Reid, “Virtual reality in pediatric neurorehabilitation: attention deficit hyperactivity disorder, autism and cerebral palsy,” Neuroepidemiology, Vol. 36, No. 1, Nov. 2010, pp. 2-18.

[14] S. Bermúdez i Badia, M. da Silva Cameirão, “The Neurorehabilitation Training Toolkit (NTT): a Novel Worldwide Accessible Motor Training Approach for at-Home Rehabilitation after Stroke” Stroke Res Treat, Vol. 2012, Article ID 802157, 13 pages, 2012. doi:10.1155/2012/802157

[15] J.H. Cauraugh, S.B. Kim, A. Duley, “Coupled bilateral movements and active neuromuscular stimulation: intralimb transfer evidence during bimanual aiming,” Neurosci Lett, Vol. 382, 2005, pp. 39-44

[16] A.H. Mason, “Coordination and control of bimanual prehension: Effects of perturbing object location,” Exp Brain Res, Vol. 188, No. 1, 2008, pp. 125-139.

[17] J.A.S. Kelso, D.L. Southard, and D. Goodman, “On the coordination of two-handed movements,” J Exp Psychol: Hum Percept Perform, Vol. 5, 1979, pp. 229-238

[18] G.M. Jackson, S.R. Jackson, and A. Kritikos, (1999). “Attention for action: Coordination of bimanual reach-to-grasp movements,” Brit J Psychol, Vol. 90, 1999, pp. 247–270.

[19] G.P. Bingham, K. Hughes, and M. Mon-Williams, “The coordination patterns observed when the hands reach-to-grasp separate objects,” Exp Brain Res, Vol. 184, 2008, pp. 283–293.

[20] J. Diedrichsen, R. Nambisan, S. Kennerley, R.B. Ivry, “Independent on-line control of the two hands during bimanual reaching,” Eur J Neurosci, Vol 19, 2004, pp. 1643–1652.

[21] P.J. Sparto, M.S. Redfern, J.G. Jasko, M.L. Casselbrant, E.M. Mandel, J.M. Furman, “The influence of dynamic visual cues for postural control in children aged 7-12 years,” Exp Brain Res, Vol. 168, Jan. 2006, pp. 505-516.

[22] S. Chaput, L. Proteau, “Aging and motor control,” J Gerontol B Psychol Sci Soc Sci, Vol. 51, Nov. 1996, pp. 346-355.

[23] S. Zoia, E. Pezzetta, L. Blason, A. Scabar, M. Carrozzi, M. Bulgheroni, and U. Castiello, “A comparison of the reach-to-grasp movements between children and adults: a kinematic study,” Dev Neuropsychol, Vol. 30, 2006, pp 719–738.

[24] G.E. Stelmach, P.C. Amrein, N.L. Goggin, “Age differences in bimanual coordination,” J Gerontol, Vol. 43, Jan 1988, pp 18-23.

[25] A.H. Mason, J.L. Bruyn, J.C. Lazarus, “Bimanual coordination in children: Manipulation of object distance”, J. Sport Exer Pscyhol, Vol. 34, suppl., 2012, p. s110.

[26] P.J. Grabowski and A.H. Mason, “Vision for Motor Performance in Virtual Environments Across the Lifespan,” In Virtual Reality and Environments, Cecilia Sik Lanyi, Ed. Croatia: InTech Press, 2012, ISBN: 978-953-51-0579-4.

[27] R.A. Magill (2011) Motor Learning and control: Concepts and Applications (9th edition), New York, McGraw-Hill.

Usability Analysis of Children's iPad Electronic Picture books

Pei-shiuan Tsai
Department of Early Childhood Education
Taipei Municipal University of Education
Taipei, Taiwan
email: peishian@gmail.com

Manlai You
Graduate School of Design
National Yunlin University of Science and Technology
Yunlin, Taiwan
email: youm@yuntech.edu.tw

***Abstract*—The main purpose of the research is to understand the current situation of design and development of iPad electronic picture books and analyze the usability. The researcher used the ranking lists search and browsing in the Apple Store for browsing various electronic picture books in great number. In the final stage, we screened out six different models to be used in iPad electronic picture books and conducted the analyses of usability. We selected by purposive sampling 15 adults (including eight teachers and seven mothers) and six six-year-old children (3 boys and 3 girls) who had experiences of using iPad. The subjects at first browsed six electronic picture books. Then 15 adults filled out the questionnaires, six children were interviewed and their operations were observed to understand their preferences and the uses of the products. The research found that the commonalities of the design of iPad electronic picture books a) dominated by page style; b) focused on linear development; c) most of interactive designs of story contents are clicking the objects on the screen; d) provided different languages and audio versions; e) most of them adopted the design of limited animation in which they used zoom-in and zoom-out and diversion techniques to show animations. The recommendations for future publishers and designers were: a) increase the interactions of story content; b) increase traditional Chinese subtitles and voice; c) integrate picture book platforms.**

Keywords- Usability; e-Picture Book; iPad Picture book

I. INTRODUCTION

The 3C products such as tablet PC, iPad, smart phone and e-book reader are unknowing invade our daily life. For the new-generation of children, the time and opportunities for reading paper books hence become less and less while the time spending on 3C products are more and more. Therefore, the contents provided by those tools become a focus point of concern. The new reading equipments not necessarily enable the children to love learning and reading, but undeniably, equipped by e-book reader and combined with multimedia elements, the reading contents have provided a different reading experience. When the iPad launched in May 2010, it was reported to be one of the most popular electronic devices [1][2]. iPad, combined a touch screen and multimedia, provides an experience of more intuitive operation than web pages and CD versions. The main purpose of the research is to understand the current situation of design and development of iPad electronic picture books and analyze their usability.

II. LITERATURE REVIEW

A. Electronic Book

The Chinese term “electronic book” is directly translated from English. In literature, Van Dam mentioned electronic books for the first time. In a broad sense, it means the media that stores and transmits the characters and pictures information through electronic channel [3]. Barker [4] argued that the electronic book was used to describe new type of books that was different from traditional paper books. But like paper books, they were composed of pages. The difference was that each page of an electronic book was designed and dynamic electronic information. Electronic book could be considered an aggregation of multi-pages, responsive and lively multi-media (includes information of characters, picture or voice). A picture book is an art form that combines visual and verbal narratives in a book format. A true picture book tells the story both with words and pictures. Electronic picture book (or e-Picture book, EPB) is to present picture books in the electronic form including CD-ROM, WWW. The applied multi-media elements include characters, pictures, animations, voice, sound effects and music. It mainly operates through mouse and keyboard in user control (interactive operation pattern). The manipulation of mouse includes drag and click whereas the manipulation of the keyboard I include character enter and key enter. The source of story materials includes adaptation and creation. The E-book in the research means iPad e-Picture books.

B. Usability

Schneiderman [5] emphasizes consistency and predictability in interface design that provides for a high level of user control. Usability means that the people who use the product can do so quickly and easily to accomplish their own tasks. This definition rests on four points: (1) Usability means focusing on users; (2) people use products to be productive; (3) are busy people trying to accomplish tasks; and (4) users decide when a product is easy to use [6]. Lazar [7] highlights ease-of-use as an equally important usability consideration he also advocates for a balanced

 ISBN: 978-1-61208-250-9

approach to Web design that allows for the appropriate use of media elements such as graphics, plug-ins, and animation.

Usability is the quality of attribute that assesses how easy user interfaces are to use. The word "usability" also refers to methods for improving ease-of-use during the design process. Usability is defined by five quality components: (1) Learnability: How easy is it for users to accomplish basic tasks the first time they encounter the design? (2) Efficiency: Once users have learned the design, how quickly can they perform tasks? (3) Memorability: When the users return to the design after a period of not using it, how easily can they reestablish proficiency? (4) Errors: How many errors do users make, how severe are these errors, and how easily can they recover from the errors? (5) Satisfaction: How pleasant is it to use the design? [8]

In conclusion, usability includes considerations such as: (1) Who are the users, what do they know, and what can they learn? (2) What do users want or need to do? (3) What is the general background of the users? (4) What is the context in which the user is working? (5) What has to be left on the machine? Usability is the ease of use and learnability of a human-made object.

III. Methodology

A. Procedure

1. In January 1 to January 31, 2011, the researchers searched for electronic picture books in the Books category of Apple App Store, and found 65 production companies publishing electronic picture books for iPad in total. According to the overall design, 6 more distinctive electronic picture books were chosen by the researchers for further questionnaire, interviews and observation.
2. The iPad 2 (Wi-Fi only) with 64GB memory and iOS 4.3 was selected for the study.
3. A questionnaire survey of five-point Likert items (1: strongly disagree, 2: disagree, 3: neither agree nor disagree, 4: agree, 5: strongly agree) and interviews were performed to adult users. The content of the questionnaire mainly included overall design, easy operation, story animation design, text design, and voice design. The interviews mainly involved questions about the operations of electronic picture books from different websites, and finding out the reasons of users' satisfaction or dissatisfaction.
4. Observations of child users' operations and interviews with those users were made. The researchers observed child users' behaviors in the operation process, and interviewed them about the issues they had encountered in their operations.

B. Study Subjects

1. 15 adults who had not used the iPad before (4 elementary school teachers, 4 kindergarten teachers and 7 mothers who had at least one child) randomly browsed the 6 electronic picture books, then filled out the questionnaire. 8 of the adults had master's degrees, and 7 had bachelor's degrees. Every subject took about 1 hour to finish the process.
2. 6 children who had not used the iPad before (six 6-year-old, half were boys and half were girls) randomly browsed the 6 electronic picture books, and were observed and interviewed by the researchers. Every time after finishing one electronic picture book, each child was asked if the rest was needed, and took rests when necessary. Every subject took about 1-1.5 hours to finish the process.

IV. Results and Conclusions

The results of the questionnaire and interviews are summed up as follows:

A. Overall Design

Although the styles of the six electronic picture books were different, users all had pretty good satisfaction to them. The scores the picture books got on the questionnaires filled by the adults were all 4.1 and above. The children liked the picture books as well. For example, C3 said "I hope Father and Mother will buy them for me, I like every one of them.", C5 said "It would be great if our textbooks were this interesting."

B. Operation Design

The user can operate the flip function of electronic picture books. The picture books that had left and right arrow marks on the screen were instantly understood by users. Children needed the researchers' reminding to roll over pages, but had no difficulties on operation either. Also, they became more skilled in operating auto and manual play, and text / voice switch, after they encountered them twice. Child users said that it was really convenient to click with fingers. "Snow White - 3D Pop-up Book" had the most different design among the six electronic picture books, but could as well be smoothly operated with hints given. The recording function provided by some of the picture books was very fresh to the subjects. For example, the child C2 said "It is fun that you can record the story you read."

C. Text Design

For the children in Taiwan, Chinese is their native language. Electronic picture books that were presented only in English were still difficult for the subjects. Adult subjects applauded that electronic picture books provided versions of multiple languages for operators. For example, the teacher T4 said "Take Little Snail as an example. It provides many different text languages so that more people in the world can browse it. That is what present multimedia design can do." Only most adult subjects thought that the traditional Chinese version should be modified for Taiwan's readers.

D. Illustration Design

The six electronic picture books have different styles and each subject has his own preference. As a whole, the average scores of adult questionnaires were over 3.9. "I love all of them, but 'The Three Little Pigs' is the one I love most because the pig in the book can move; 'Snow White' is also very special and I have never seen such book."

E. Voice Design

Text and voice is mutually collocated that most of the adult subjects for the part of the Chinese voice proposed that it be appropriate to find native Taiwanese for dubbing to avoid the interference from an accent in listening. The adult subjects affirmed that the English text was helpful to the non-English speaking readers for the enhancement in learning English, but proposed to add the Chinese that more Taiwanese audience could participate in.

F. Animation Design

Most of the electronic picture books on the current iPad are presented without or with limited animation. The limited animation is dominated with the movement of leading characters or part of objects, or zoom-in and out and movement of camera shots. The child subjects showed high interest in the dynamic performance. When the researcher hinted that some figures or objects in the frame can be clicked with fingers. For children, it is a very new try and they will try to click to see if there is any reaction.

As a whole, there are two primary common points in the six electronic picture books: 1) single-line development of the story; 2) the contents of the story lacks of interactions that both adult and child users indicated that they wish to read more electronic picture books if they have a chance because there are many differences comparing to reading physical books.

V. SUGGESTIONS

This research evaluated the operation and uses of six selected electronic picture books and gave following suggestions for the reactions from the subjects. It is hoped that in future, more selected picture books and subjects could be used for the usability evaluation that will prompt more concrete contributions to the designs of electronic picture books.

A. Enhance the interaction of the story

Given the fact that many picture books were adapted from printing picture books in which the story is developed in single-line, the interaction of story contents is limited. The advantages of multi-media are that it could facilitate the increase of interaction. In the future, we could bring the characteristics into full play by increasing double-line or even multi-line developments of the story and the design of interaction with the contents of the story to enhance the interaction between readers and story contents.

B. Increase Traditional Chinese Subtitles and Voice

At the present time, only a few developers provide traditional subtitles and voice in their electronic picture books including Rye Studio、Apple Style、RYBooks Studio. Many other excellent electronic picture books choose English as their only or primary language. But they can just add traditional Chinese subtitles and voice to meet the requirement from readers in Taiwan. So we suggest that the developers cooperate with foreign developers in adding traditional Chinese subtitles and voice for the readers in Taiwan that more children have a chance to read the rich contents of electronic picture books.

C. Integrate Picture book Platforms

Although there is classification of books in Apple App Store, searching for specific electronic picture books is a time-consuming job. Therefore, we suggest that create a classification for the electronic picture books or design a browser search interface exclusively for the children.

In summary, from the angle of emotional interaction, operation interaction, cognition interaction and community interaction, the design of existing electronic picture books could strengthen the cognition interaction and community interaction that the visitors could interact with the contents of electronic picture books through the characteristics of multi-media factors that are different from the printing books. Also, it could have meaningful community interactions with other readers through the linking of Internet. With the creativity and innovation of interactive technology, more and more excellent iPad picture book worthy of researchers continue to invest in research.

REFERENCES

[1] H. Anderson, iPad: The savior of digital publishing. Journal of Internet Law, 2011, 14(10), pp.15–20.

[2] O. Murray and N. Olcese, Teaching and learning with iPads, reading or not? TechTrends, 2011, 55(5), pp.42–48.

[3] Lwan-Syin. Lwo, Electronic books and the new communication era. Instructional technology and media, 1995, 21, pp.13-16. (Chinese)

[4] P. Barker, Electronic books and libraries of the future, The Electronic Library, 1992, 10(3), pp.139-149.

[5] B. Schneiderman, Leonardo's laptop: Human needs and the new computing technologies. Cambridge: The MIT Press, 2002.

[6] J. S. Dumas and J. C. Redish, A practical guide to usability testing. Portland, OR: Intellectual books, 1999.

[7] J. Lazar, Web usability: A user-centered design approach. Boston: Pearson Education, Inc., 2006.

[8] J. Nielsen , Usability 101: Introduction to Usability, 2003, Retrieved December 20, 2008 from http://www.useit.com/alertbox/20030825.html

 ISBN: 978-1-61208-250-9

TABLE I. THE BASIC INFORMATION OF THE SIX ELECTRONIC PICTUREBOOKS

Story Name	The Little Snail	One Pizza, One Penny	Just Grandma and Me	Bedtime Monster	The Three Little Pigs	Snow White - 3D Pop-up Book
Publisher	Rye Studio	Apple Tree & Guru Bear	Oceanhouse Media	Siena Entertainment, LLC	Nosy Crow	lee hee suck
Illustration Design	2D	2D	2D	2D	3D	3D
Language (Text)	English, Traditional Chinese, Simplified Chinese, Japanese, French, German, Spanish	English, Traditional Chinese, Simplified Chinese	English	English Spanish	English	English
Language (Voice)	English, Chinese, Japanese, French, German, Spanish	English, Chinese	English	English Spanish	English	English
Sleeping Mode (voice only)	✓	✓	×	×	×	×
Record	✓	✓	×	×	×	×
Flip Mode	1.Auto 2.Manual (click on last/next page button to turn)	1. Auto 2.Manual (click on bottom left/right corner to turn)	1. Auto 2. Manual (roll over the page to turn)	1. Auto 2.Manual (click on last/next page button to turn)	1. Auto 2.Manual (click on last/next page button to turn)	1. Auto 2.Manual (click on last/next page button to turn)
Page Index	×	×	×	✓	×	×
Bookmark	×	✓	✓	✓	×	×
Operating Instructions	✓	×	×	✓	✓	×
Story Scenes Content Clicking	×	×	✓	×	✓	✓
Simple Animation	×	×	✓	×	✓	×
Extended Activities	×	✓	×	×	×	✓

Figure 1. The Little Snail

Figure 2. One Pizza, One Penny

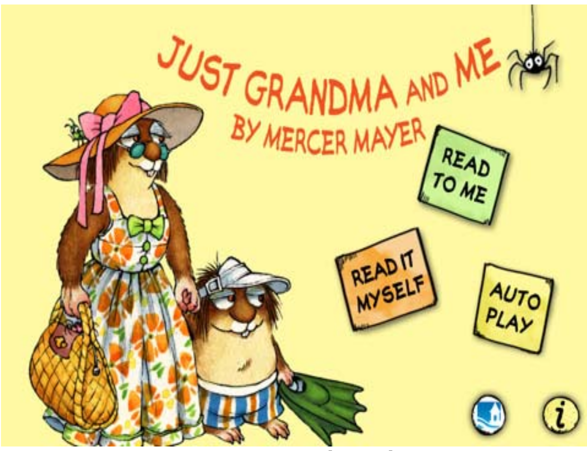


Figure 3. Just Grandma and Me

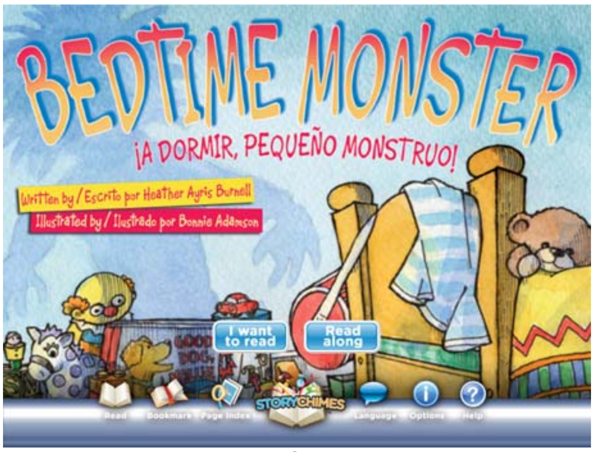


Figure 4. Bedtime Monster

Figure 5. The Little Snail

Figure 6. Snow White - 3D Pop-up Book

Evaluating the Impact of Spatial Ability in Virtual and Real World Environments

Georgi Batinov, Kofi Whitney, Les Miller, Sarah Nusser and Bryan Stanfill
Iowa State University
Ames, Iowa, USA
{batinov,kwhitney,lmiller,nusser,stanfill}@iastate.edu

Kathleen T. Ashenfelter
US Census Bureau
Washington, D.C. 20233
kathleen.t.ashenfelter@census.gov

***Abstract*— Survey agencies in the United States continue to move many map-based surveys from paper to handheld computers. With large highly diverse workforces, it is necessary to test software with a diverse population. The present work examines the performance of participants grouped by their level of spatial visualization. The participants were tested in either the field or in a fully immersive virtual environment. The methodology of the study is explained. The performance of the participants in the two environments is modeled with least squares regression. Results of the study are presented and discussed.**

Keywords- map-based survey; virtual reality; spatial ability

I. INTRODUCTION

Survey agencies in the United States have been moving towards using handheld computers to replace the use of paper in their field operations. Since most field surveys are inherently location dependent, a lot of the software used will be map-based. Agencies, like the Bureau of Census, are forced to couple this move to map-based software with a highly diverse workforce, especially in large scale operations like the decennial census. Due to the wide range of individual differences typically encountered in such diverse workforces, software testing is a critical component of this process. Ultimately, the software has to be tested in the field to fully understand how it will perform. However, a significant issue with testing in the field is the cost. An interesting question is the viability of doing at least the initial testing of software in a virtual environment.

In the present work a study is described that looks at the participant's performance in either the field or in a fully immersive virtual environment. The task chosen for the study was address verification, where a census worker is given a list of addresses and they are expected to either determine the address is correctly located on their map or make the necessary corrections. The contribution of this paper is the direct comparison of a complex real world operation performed in both the field and virtual reality. This initial study didn't show very much difference between performance in the field and virtual environments. We did see a significant impact of the role that spatial visualization played in both environments.

In the remainder of the paper we look at related work, examine the methodology used in the study, present the statistical results, and provide a discussion of what we found.

II. BACKGROUND

Wobbrock et al. [24] proposed ability-based design as a paradigm for constructing individual-centric systems. According to Murray & Kluckhohn [18],"Every man is in certain respects (a) like all other men, (b) like some other men, (c) like no other man" (p.35). Benyon, Crerar, & Wilkinson [3] predicated the prominence of cognitive differences in human-computer interaction on the divide between physical and digital artifacts and noted that cognitive differences may have amplified effects in computing contexts (pp. 21-22).

Spatial ability is a compound factor that has often been linked to performance in interactive tasks. Several authors have used factor-analytic techniques to decompose spatial ability into constituents. In the nineties, [5] and [16] reported that it consists of five parts: spatial visualization, speeded rotation, closure speed, closure flexibility, and perceptual speed. Earlier publications by [6] and [20] suggested other combinations. Spatial visualization ability–defined by [9] (p. 173) as "the ability to manipulate or transform the image of spatial patterns into other arrangements"–has been shown to correlate with performance with command-line interfaces [11, 4], file system navigation [22], searching an information retrieval system [8], web browsing [25], simulated driving [1], and remote control of robots [15].

Beyond the combination of software and hardware, we also need to consider the field operating environment, which presents a multitude of stimuli and a continuously changing external context, unlike the traditional computer desktop. Whether a field setting can be reasonably approximated in a laboratory virtual environment is still an open question. Ref. [17] highlights the tradeoff between experimental control and ecological validity in traditional research methods and suggest that improved-fidelity virtual reality may reduce the compromise. If correct, their claim has practical implications, as well: virtual reality may become a low-cost alternative for field training. Two components that distinguish reality from a virtual environment are distance perception [23] and embodiment [2]. Ref. [10] compared environmental learning from the real world, non-interactive video, and a desktop virtual environment, concluding that spatial ability is correlated with learning in both the real world and virtual environment, with a stronger effect for the desktop simulation. Ref. [21] compared walking patterns in reality and on a treadmill in non-immersive virtual setting, noting persisting small differences in gait after 20 minutes of acclimatization. Ref. [19] showed that increasing display size and resolution improved wayfinding and object location performance in a non-immersive virtual setting. The outcomes in these publications suggest that as we improve visualization and locomotion technologies, we may be able to run virtual reality experiments approaching ecological validity. Refs. [13] and [14] used a study design that appears similar to ours but was driven by a different agenda. The authors constructed a high-fidelity virtual reality model of a residential area in the United Kingdom and asked 27 participants to navigate to five locations inside an immersive environment designed by [7]. Participants had access to schematic maps, detailed maps, and written and spoken route instructions on a handheld device that served as a pathfinding aid. The authors described three different pathfinding behaviors, noted geographical hot spots for handheld device activity and per-destination aggregate device activity. The focus of the research was on linking location to handheld usage, and the authors did not report measures of statistical validity. In contrast, we set out to find statistical evidence for performance differentials on a software map task. Our experiment includes both a virtual reality and a field setting. Additionally, our task is a software task that has a navigation component. Finally, the virtual environment in [13, 14] had a lesser degree of immersion, because participants navigated with a joystick and the virtual model was projected on up to three VR walls.

III. MATERIALS AND METHOD

The experiment contained a screening phase and an exercise phase.

A. Screening phase

During the screening phase, one-hundred-and-twenty-four participants were individually assessed on spatial visualization, visual memory, perceptual speed, and perspective-taking ability. The tests were VZ-2, MV-2, and P-2 given in [9], and the perspective-taking assessment described in [12]. Participants with spatial visualization scores greater than or equal to 15 or less than 9 (out of 20) were randomly assigned to one of two treatments in the exercise phase. Pairs from either the low or high spatial visualization groups were randomized together, allowing each participant a 0.5 probability of assignment to either be tested in the virtual reality treatment or the field treatment. Thirty-two participants (14 males, 18 females) were assigned to the second phase of the experiment.

B. Field phase

For the field treatment, 15 participants (8 males, 7 females) were taken individually to the same spot in a residential neighborhood in Ames, Iowa. They were first trained on using the handheld device, locating addresses in the field, and the think-aloud protocol. An observer provided them with a stylus and a handheld computer: a Pharos Traveler 535x with a 240x320, 3.5” transflective screen and a 624 MHz Intel PXA270 processor. The observer explained the address verification task.

Participants would have to physically walk to an address in order to answer the question. If the map contained errors, they had to use the software’s editing features to position the address at the correct location or remove it altogether. Four outcomes were possible. An address needed to either be added to the map, deleted, moved to a new location, or confirmed without changing the map. Participants were told to only correct the addresses in their task list and to ignore other possible errors on the map. (The map contained no errors outside of scenario addresses.) Participants were then taught how to edit the software map and were also instructed to verbalize all their thoughts for a think-aloud protocol. The map software was started in training mode and participants were asked to locate and verify three training addresses in the immediate vicinity, while the observers answered procedure questions and provided feedback on the quality of the think-aloud. At the end of the training session, observers answered the participant’s final questions, and also explained that observers would not talk during the actual exercise, other than to prompt the participant to keep verbalizing or to ask about behavioral details. Observers then returned the participant to the location where all trainees started, switched the map software to experiment mode, and started an audio recorder (worn by the participant) and a GPS tracker (carried by the observer). The GPS tracker was not given to the participant so that they would not be interrupted to time-stamp address completions.

All participants verified the same six addresses off of an identical randomized list order (Figure 2), and therefore could not benefit from completion sequence hints. The list could be viewed at all times in the software by tapping the currently selected address. Errors for each address in the task list are shown in Figure 1. Participants were allowed to work on addresses in any order and could return to previously submitted addresses as many times as they wanted. Only final answers were evaluated for correctness. When finished, participants were then taken to two locations on the map and asked to point in the direction of the starting spot. Finally, observers audio-recorded an exit questionnaire detailing the participant’s perceptions of the study.

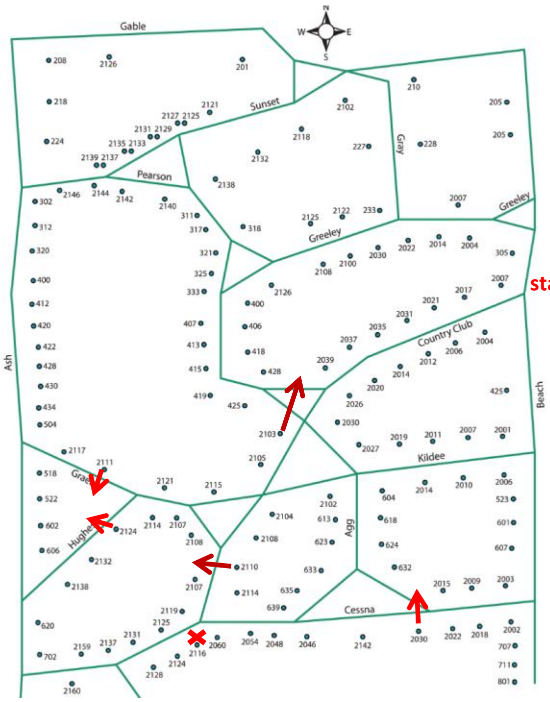


Figure 1. Address errors introduced on the map.

C. Virtual reality phase

Seventeen participants (6 males, 11 females) were randomly assigned to the virtual reality treatment and were taken individually to the VRAC C6 immersive virtual reality environment on the Iowa State University campus.

Virtual reality model-The virtual setting loaded in the environment was a high-fidelity three-dimensional model of the residential area portrayed in Figure 1, with an additional block modeled outside the westernmost and easternmost extents of the map. The model was created in SketchUp (www.sketchup.com) and imported into the virtual reality environment through VR Juggler (www.vrjuggler.org). Housing units and streets were georeferenced. However, actual housing units were represented by house models of similar size and style selected from Sketchup's repository of three-dimensional housing models (sketchup.google.com/ 3dwarehouse/). The neighborhood model also incorporated notable landmarks in the area, such as, street signs, curbs, textured surfaces, a day sky with sun, trees, shrubs, a playground, and a large building on the Iowa State University campus that was visible in some parts of the study area. The model did not include sidewalks, but did represent multi-lane streets and split boulevards, keeping throughway widths consistent with reality.

Virtual reality equipment-The virtual reality room is a cube with dimensions 3.05 x 3.05 x 3.05 m. Each of the four walls, floor, and ceiling displayed stereo images of 4096 x 4096 pixels at approximately 16 frames per second. Video projection is driven by a cluster of 48 HP xw9300 workstations with 96 nVidia Quadro graphics cards sending video frames to 24 Sony SRX-S105 digital cinema projectors. InterSense's IS-900 tracking system tracked the participant's head location and gaze direction, and the stereo perspective dynamically shifted with the user's gaze. The participants wore active stereo glasses.

Moving in virtual reality-Movement in the environment was accomplished by stepping towards the desired direction. A circular spot in the center of the floor, approximately 0.6 m in diameter, was the "dead zone". If the participant's head was located in the column of the spot, all movement stopped. Stepping outside the dead zone would start moving the virtual reality model in the opposite direction of the step, giving the illusion of the participant moving through the model in the direction of the step. As the participant stepped closer to the walls, movement speed increased, from approximately 0.1 m/s to a maximum of approximately 2.22 m/s (8 km/h or 5 mi/h). We fixed the maximum speed to a slow trot, because we were concerned that a higher speed could not be encountered in the range of walking speeds available to participants in the field treatment, and a lower maximum speed might bore participants, causing them to lose focus.

Protocol changes-Study protocol was exactly the same as in the field treatment, but prior to introducing the handheld device, participants were trained on moving inside the virtual environment. Participants also started training and the exercise at the same geographic spot in the virtual model as participants in the field.

Data collection and analysis-We tracked: distance traveled via GPS and virtual movement logs, time taken to complete the task, and number of addresses incorrectly verified (number of task errors). The virtual reality model was georeferenced, so travel coordinates within the immersive environment reflected actual distances. Additionally, we recorded all handheld software actions and user speech from the end of the training session to the end of the exit questionnaire.

We used least squares regression to explore statistical relationships among the data. Our response variables were distance traveled, time taken to complete the task, and number of errors. Predictor variables included spatial visualization category (low or high); field/virtual environment category; gender; perceptual speed, visual memory, and perspective-taking scores; and zoom, pan and map reset actions.

We hypothesized that:

Hypothesis 1: High-spatial-visualization participants would travel significantly shorter distances than low-spatial-visualization participants in both the field and virtual environments.

Hypothesis 2: High-spatial-visualization participants would take significantly less time than low-spatial-visualization participants in both the field and virtual environments.

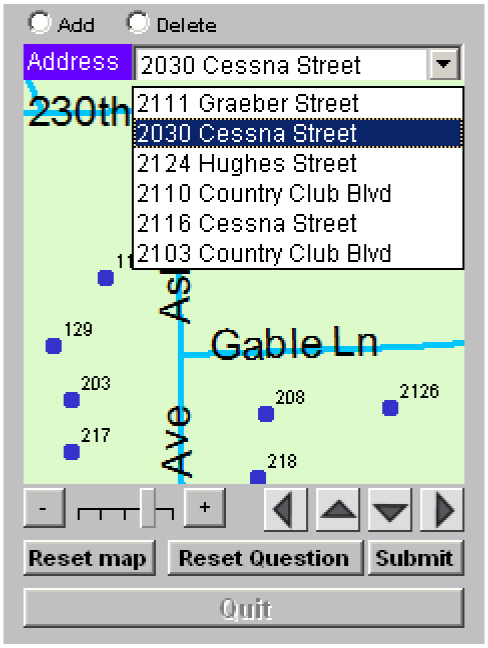


Figure 2. Edit screen with address list extended.

Figure 3. Participant in the virtual reality treatment.

In addition, we expected that there would be some impact on the participants between treatments, especially the participants with low spatial visualization in the virtual environment.

IV. RESULTS

The results reported in this paper are based on least squares regression models. The focus has been to look at two slightly different sets of variables. The first model looks at how a set of variables based on the environment and the participant impact performance measures (e.g., time and distance), while the second model adds a software flavor to the analysis by adding a variable that incorporates map resets and pans.

A. Regression based on Environment and Participants

The first model examines the impact of the environment (E = 0 Field or 1 Virtual), spatial ability (S = 0 High Visualization or 1 Low Visualization), and gender (G = 0 Female or 1 Male) on time and distance.

In particular we looked at Y = E + S + E*S + G, where Y is the prediction of either log(time) or log(distance) and E*S is the interaction of the environment and the participant's spatial ability assignment.

The regression results for log(time) and log(distance) are shown in Tables I and II, respectively.

TABLE I. REGRESSION RESULTS FOR LOG(TIME).

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	3.338	0.112	29.814	0.000
Env	0.083	0.159	0.524	0.605
Spatial	0.398	0.152	2.622	0.014
Gender	0.005	0.107	0.050	0.961
Env:Spatial	0.014	0.209	0.068	0.946

TABLE II. REGRESSION RESULTS FOR LOG(DISTANCE).

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.021	0.117	0.177	0.861
Env	0.268	0.166	1.617	0.117
Spatial	0.373	0.159	2.355	0.026
Gender	0.029	0.111	0.261	0.796
Env:Spatial	-0.425	0.218	-1.944	0.062

The most interesting aspect of the results shown in the two tables is the significance of the participants' level of spatial ability in both results (Pr(>|t|) = 0.014 and 0.026, respectively. The regression model for distance (Table II) is suggestive that the interaction of E and S is important with Pr(>|t|) = 0.062.

The box plots for log(Time) and log(Distance) mediated by the interaction are shown in Figures 4 and 5, respectively. While the box plots don't provide too much information, they do provide some insight. First, in Figure 5 it appears that high spatial participants in the virtual environment traveled greater distances than the high spatial participants did in the field. The log(Time) boxplot (Figure 4) is suggestive that spatial ability is important in terms of

 ISBN: 978-1-61208-250-9

the time taken (log(Time)). Such a result makes sense in light of the significance of spatial ability seen in Table I.

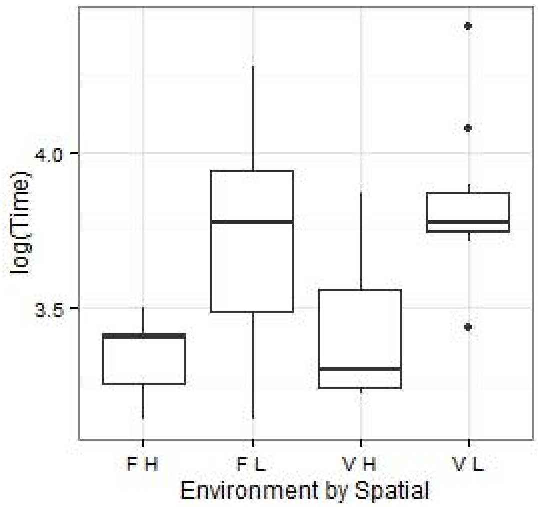


Figure 4. Boxplots for log(Time) vs E*S.

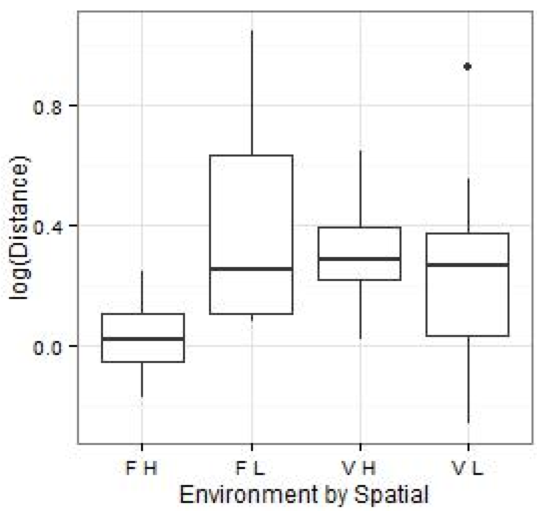


Figure 5. Boxplots for log(Distance) vs E*S.

B. Including Software Operations in the Regression

To bring the software performance of the participants into the regression model, we defined a variable (Resetpan) as a measure of how many times a participant interacted with the map on the handheld device. Formally, ResetPan is defined by

Resetpan=0.5(Resets-mean(Resets)/sd(Resets) + Pans-mean(Pans)/sd(Pans)).

The choice of map resets and pans was motivated by the experiment observers' experience in both the field and virtual environments. Tables III, IV, and V show the regression results for the least squares regression models for log(Time), log(Distance), and Errors, respectively.

TABLE III. REGRESSION RESULTS FOR LOG(TIME).

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	3.111	0.12	25.979	0.000
Env	0.041	0.138	0.297	0.769
Spatial	0.301	0.134	2.235	0.034
Gender	0.039	0.093	0.424	0.675
Resetpans	0.036	0.011	3.213	0.003
Env:Spatial	-0.005	0.181	-0.026	0.979

TABLE IV. REGRESSION RESULTS FOR LOG(DISTANCE).

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	-0.200	0.128	-1.563	0.130
Env	0.227	0.148	1.541	0.135
Spatial	0.278	0.144	1.932	0.064
Gender	0.062	0.099	0.626	0.537
Resetpans	0.035	0.012	2.915	0.007
Env:Spatial	-0.443	0.193	-2.292	0.030

TABLE V. REGRESSION RESULTS FOR ERRORS.

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	2.219	0.772	2.875	0.008
Env	-0.123	0.888	-0.138	0.891
Spatial	1.811	0.867	2.089	0.047
Gender	0.338	0.597	0.565	0.577
Resetpans	-0.029	0.073	0.000	0.696
Env:Spatial	-0.919	1.164	-0.790	0.437

From the results in the three tables, it appears that the spatial ability levels of the participants was again an important variable as it is significant for log(Time) (Pr(>|t|) = 0.034 in Table III) and for Errors (Pr(>|t|) = 0.045 in Table V). It was also suggestive for log(Distance) (Pr(>|t|) = 0.064 in Table IV). As expected, Resetpans showed up as significant for both log(Time) and log(Distance) results. Interestingly, it doesn't show up as a factor in the number of errors made by the participants. Finally, as in Table II, the interaction between the environment and the level of spatial ability was significant for log(Distance) (Pr(>|t|) = 0.031 in Table IV).

V. DISCUSSION

The spatial visualization ability level of the participants was a significant factor in the regression models for all but the regression model for log(Distance) using the Resetpan variable and was suggestive there. The boxplots in Figures 4 and 5 show some support for the hypotheses given earlier except for the virtual environment participants in the log(Distance) boxplot. Somewhat surprising was how little impact the environment (and the interaction between the environment and spatial ability) had in the experiment. Beyond the suggestion of a difference in the distance traveled between high spatial visualization ability people in the field versus high spatial visualization ability people in the virtual environment (Figure 5), we did not find a separation based on the environment.

Two issues that made it difficult to work with the data were the number of participants used in the study and the high level of variation that we found in the performance of low spatial participants, especially in the virtual environment. The number of participants that we used in the experiment was a function of the difficulty that we had in finding participants with a low level of spatial ability. In spite of testing a large pool of subjects in Phase I, we struggled to find a sufficient number of low spatial ability participants to increase the size of the experiment.

VI. CONCLUSION AND FUTURE WORK

The experiments provided support for the development of map-based user interfaces that work with both high and low levels of spatial ability. Since we were unable to find a significant difference due to environment, testing software in the virtual environment remains a realistic possibility. However, we expect that we will have to do more testing with an increased Phase II sample size in order to validate such an approach.

VII. REFERENCES

[1] Andrews, E. C., & Westerman, S. J. (2012). Age differences in simulated driving performance: Compensatory processes. *Accident Analysis & Prevention*, *45*(0), 660–668.

[2] Arnold, P., & Farrell, M. J. (2003). Embodiment and spatial behavior in virtual environments: comments on Durlach et al.(2000). *Presence: Teleoperators & Virtual Environments*, *12*(6), 658–662.

[3] Benyon, D., Crerar, A., & Wilkinson, S. (2001). Individual differences and inclusive design. In C. Stephanidis (Ed.), *User interfaces for all: concepts, methods, and tools*. Mahwah, NJ: Lawrence Erlbaum Associates.

[4] Benyon, D. R. (1993). Accommodating individual differences through an adaptive user interface. In M. Schneider-Hufschmidt, T. Kuhme, & U. Malinowski (Eds.), *Adaptive User Interfaces: Principles and Practice* (pp. 149–165). Amsterdam: Elsevier Science.

[5] Carroll, J. (1993). *Human cognitive abilities: A survey of factor analytical studies.* Cambridge: Cambridge University Press.

[6] Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.

[7] Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. *Commun. ACM*, *35*(6), 64–72. doi:10.1145/129888.129892

[8] Downing, R. E., Moore, J. L., & Brown, S. W. (2005). The effects and interaction of spatial visualization and domain expertise on information seeking. *Computers in Human Behavior*, *21*(2), 195–209.

[9] Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

[10] Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*(2), 151–176.

[11] Jennings, F., Benyon, D. R., & Murray, D. M. (1991). Adapting systems to differences between individuals. *Acta Psychologica*, *78*(1-3), 243–258.

[12] Kozhevnikov, M., Motes, M., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology*, *20*, 397–417.

[13] Li, C. (2006). User preferences, information transactions and location-based services: A study of urban pedestrian wayfinding. *Computers, Environment and Urban Systems*, *30*(6), 726–740.

[14] Li, C., & Willis, K. (2006). Modeling context aware interaction for wayfinding using mobile devices. *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (pp. 97–100).

[15] Liu, A. M., Oman, C. M., Galvan, R., & Natapoff, A. (2012). Predicting space telerobotic operator training performance from human spatial ability assessment. *Acta Astronautica*, (0), -. doi:10.1016/j.actaastro.2012.04.004

[16] Lohman, D. F. (1996). Spatial ability and g. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97–116). Mahwah, NJ: Lawrence Erlbaum Associates.

[17] Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods*, *31*(4), 557–564.

[18] Murray, H. A., & Kluckhohn, C. (1948). *Personality in nature, society, and culture*. Oxford, England: Knopf.

[19] Ni, T., Bowman, D. A., & Chen, J. (2006). Increased display size and resolution improve task performance in Information-Rich Virtual Environments. *Proceedings of Graphics Interface 2006* (pp. 139–146).

[20] Royce, J. R. (1973). The conceptual framework for a multi-factor theory of individuality. In J. R. Royce (Ed.), *Multivariate Analysis and Psychological Theory*. London and New York: Academic Press.

[21] Schellenbach, M., Lövdén, M., Verrel, J., Krüger, A., & Lindenberger, U. (2010). Adult age differences in familiarization to treadmill walking within virtual environments. *Gait & posture*, *31*(3), 295–299.

[22] Vicente, K. J., Hayes, B. C., & Williges, R. C. (1987). Assaying and isolating individual differences in searching a hierarchical file system. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *29*(3), 349–359.

[23] Witmer, B. G., & Kline, P. B. (1998). Judging perceived and traversed distance in virtual environments. *Presence*, *7*(2), 144–167.

[24] Wobbrock, J. O., Kane, S. K., Gajos, K. Z., Harada, S., & Froehlich, J. (2011). Ability-Based Design: Concept, Principles and Examples. *ACM Trans. Access. Comput.*, *3*(3), 9:1–9:27. doi:10.1145/1952383.1952384

[25] Zhang, H., & Salvendy, G. (2001). The implications of visualization ability and structure preview design for Web information search tasks. *International Journal of Human-Computer Interaction*, *13*(1), 75–95.

 ISBN: 978-1-61208-250-9

Software Lifecycle Activities to Improve Security Into Medical Device Applications

Diogo C. Rispoli , Lourdes M. Brasil
Graduate in Biomedical Engineering
University of Brasília at Gama
Brasília, Brazil
e-mail: drispoli@gmail.com, lmbrasil@gmail.com

Vinicius C. Rispoli
Faculty of Engineering at Gama
University of Brasília
Brasília, Brazil
e-mail: vrispoli@unb.br

Paula G. Fernandes
Computer Science Department
University of Brasilia
Brasília, Brazil
e-mail: paulag6@gmail.com

***Abstract*—This work proposes a methodology to include into Medical Software Development Lifecycle activities that helps improve security. The methodology uses assessment techniques and methods, applied to each phase of software lifecycle, that address security concerns and help to improve software quality. As a result, a partial analysis using the methodology proposed was performed in medical software at development stage to help reduce its gap between safety and security requirements.**

Keywords-information security; security software; hackers; medical software lifecycle; security risks.

I. INTRODUCTION

In a scenario of constant technological evolution, the demand for solutions in medical field is constant. Nowadays, you can find free software available on the internet which collect vital information of individuals and can be installed on mobile devices, such as smartphones and tablet computers [1]. Despite this speed, standards focused on medical field do not present models that make possible to assess problems associated with common security vulnerabilities that may appear in the software development cycle.

Standards as ISO / IEC 62304:2007 [2] dealing with the medical equipment software development, although recent, do not handle with these new technological perspectives. On the other hand, ISO 27799:2008 [3] deals with security concerns of health information systems, but it does not address solutions related to secure software development, compared to the present moment.

This very moment of technology effervescence opens doors for hackers to exploit and promote invasions as the attack on an insulin pump documented and presented in [4]. Based on a simple technique that combines programming skills and basic electronics, the hacker undertakes an attack on an insulin pump, used by himself, in order to demonstrate the innocent perspective that these devices are built. As a final result, he can apply a lethal dosage of insulin breaking the authentication security required by the equipment wireless communication.

Remembering that this was not the first case of attack documented on medical devices. In 2008, a U.S. team of researchers published a paper that showed an attack on a pacemaker, which also exposed security flaws related to wireless equipments [5].

Evidently, there is a rush in adopting standards, and actions, to ensure the security of the software built for medical devices. The exploitation of vulnerabilities in medical equipment can lead to death or serious injury, fraud, unauthorized disclosure of information, theft, and other attacks. For this reason, it is necessary to ensure that information security requirements (integrity, confidentiality, availability and non-repudiation of data collected) are met as well as ensuring that software vulnerabilities are not included in these devices during its development lifecycle.

This article will discuss requirements for improving the security of medical applications based on risk assessment of information security and the correlation of requirements for software security standards and their mitigation techniques related to safety in life support. As a result of this work, activities, also known as touch points, will be shown and assessed through a software development lifecycle helping to ensure the security requirements needed to consider software secure and safe.

This document is divided in six sections. In the next section, the relationship between risk perspective from safety and security views will be discussed. In section three, we will present the importance of software lifecycle and the incorporation of security activities into software construction. In the fourth section, the assessed software and its characteristics will be presented. Software assessment against the methodology proposes will be shown the in fifth section. And, in last section, will be discussed the assessment results.

II. SECURITY RISKS

In order to associate issues that are seemingly disconnected, it is important to observe how software security aspects are linked to medical devices construction. Under the perspective of the software, the risks are paths through the application where attackers (hackers) can disrupt business or organizations [10].

Observing this look from the perspective of the equipment, the risk (or level of concern) is an estimate of the injury severity which equipment can inflict or allow, directly or indirectly, in a patient or operator, as a result of device failure, design flaws, or because of the device employment for its intended use [9].

 ISBN: 978-1-61208-250-9

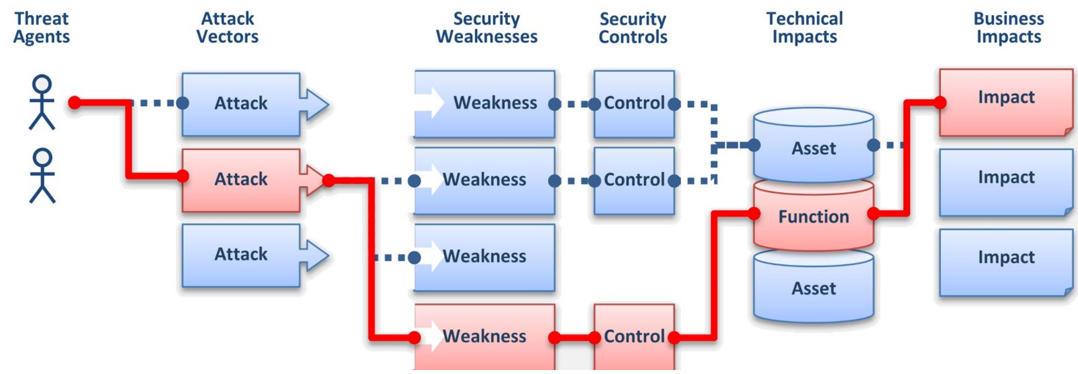


Figure 1. Applications Security Risks [10].

Objectively, these two views are very close because risks are directly related to software failures or weaknesses in its control mechanisms. Its natural consequence is the subversion and many kinds of damage, primarily damage to life, but also financial and corporate image caused by malicious people.

Threat agents can use several paths over application in order to attack organizations. These paths are through exploration of security weaknesses to bypass security controls and cause technical and business impacts, as it can be seen in Figure 1.

From this perspective, raising, mapping and balancing risks, flaws and vulnerabilities introduced by problems in the construction of medical software becomes exhausting, ineffective and away from the current technological reality. There are many patterns as you can see in Figure 2, dealing directly (such as IEC 62304:2006) or indirectly with software development lifecycle and its associated security risks [11].

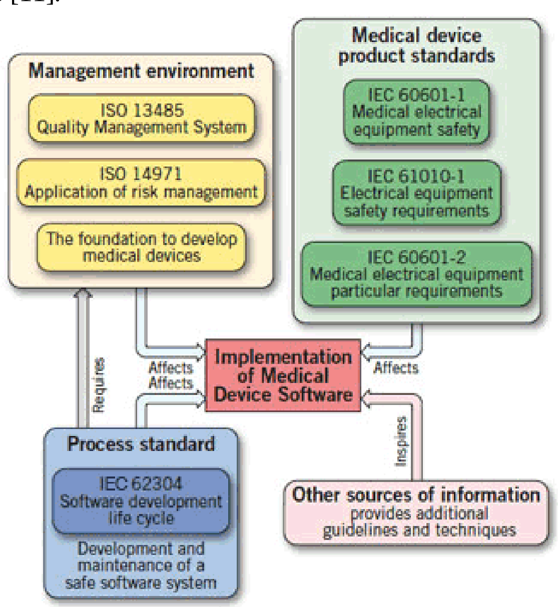


Figure 2. ISO/IEC 63304 and its relation to other standards [11].

However, there is a lack of methodologies that address mitigation aspects for exploitable vulnerabilities in software. It is important notice that IEC 62304:2006 address security as concern that manufacturers shall include in software requirements [2].

The ISO 27799:2008, which concerns to medical information systems, do not treat or address solutions related to the process of building secure software. Notwithstanding, this standard imposes requirements on the operation of informational systems as secure authentication, authorization, accountability, use of encryption, secure information communication and protection against code injection. These are relevant aspects where the software is the leading actor or an important supporting actor [3].

Observing the processes of quality assurance employed in medical applications, the aspects of validation and verification are only concerned with functional requirements of the software [10].

So, it is necessary list interactions with the lifecycle of the software that shows a path to perform penetration tests and audits, raise non-functional requirements for safe operation and deployment of applications, including risk analysis of vulnerabilities in software design, protect applications against command injection flaws and buffer overflow, properly handle errors and exceptions and logging sanitized records (after removing sensitive information) of users activity in the equipment operation [3, 8, 10, 12].

It is also important to design efficient mechanisms for authentication, secure session management, user authorization, authenticated encryption for secure transmission, storage of collected data and records of patients with the goal of increasing the guarantee of the safety and quality of information systems health and its related applications, whose assets are devices and their associated software [3, 10].

III. SOFTWARE DEVELOPMENT LIFECYCLE AND SECURITY CONCERNS

Lifecycle models organize development software activities and provide a framework to monitor and control a building software project and its future operation. Without a model is difficult to say the exact moment of project's development or validation phase and how or which situations control activities must be applied [7, 8].

Despite *Quality Systems Regulations (QSRs)* do not establish a specific lifecycle model for medical software, regulatory standards state that a model adoption is important and it should contain at least some phases like quality planning, requirements management, software project specification, coding, testing, installation, operation, support and maintenance [7, 9].

Development of checklists with controls to be applied can aid incorporation of secure coding practices throughout the construction of medical software. The use of security software techniques does not necessarily increase the cost of its development lifecycle, in order to correct problems and failures of this nature cost more after application development finishes [8].

Adoption of security techniques is expected since medical software is able to run into smartphones and other mobile devices, for example, and all information collected and transmitted by those devices are sensitive and confidential.

Software development lifecycle is part of project controls and these controls are needed to reduce flaws insertion in medical device [7]. So, to help mitigate medical software vulnerabilities problem is essential to indicate what activities must be implemented between software lifecycle phases. These activities are related to the identification, development and validation of techniques that difficult vulnerabilities exploitation in software operation.

An interesting way to improve software security and quality is perform security activities through software lifecycle. Those activities are responsible to manage security concerns and must be applied inside lifecycle phases instead deal with security concerns only at requirement phase, as suggested by IEC 62304:2006 [2]. A correlation between phases and activities can be seen in Figure 3; they were described for general software projects in [8].

It is important to notice that there is no specific methodology to use the described security touch points. They can be applied in every kind of software development lifecycle methodology [8].

A. A brief description of each touch point

The touch points are described as follows [8]:

Abuse Cases – Build abuse cases is relevant to do relationship between problems and risk analysis. At that moment is important observe if some attack pattern fits the system or software requirements. This is a good moment to model vulnerability scenarios that could be exploited in Code Review Phase and Penetration Testing Phase.

Security Requirements – Security requirements must cover functional security, safety requirements, raised abuse cases and attack patterns. In that phase every software security necessity must be mapped to ensure the correct implementation. A good example for security requirements is the correct use of cryptography to protect critical data.

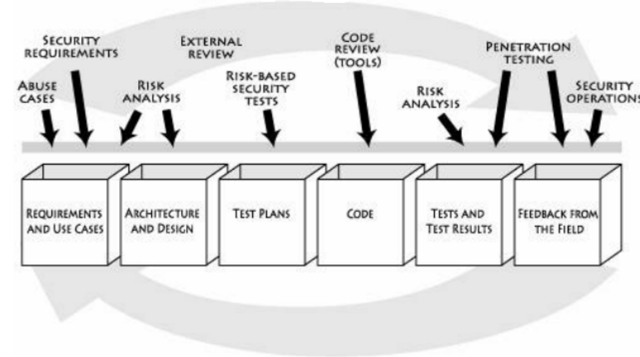


Figure 3. Security touch points inside a lifecycle [8].

Architectural Risk Analysis – Completing risk analysis oriented by ISO 14971. This analysis is a small part of a Risk Management Process that every Manufacturer must apply complying with ISO 14971, according to [2].

Risk Based Security Tests – The testing strategy must cover at least to major topics: test security requirements with standard functional testing techniques and risk-based security testing build from abuses cases and attack patterns.

Code Review – After codification, and before testing phase, the code review analysis is a good activity to ensure the security requirements were well implemented and the vulnerabilities listed in abuses cases analysis are outside the software. The code review can be automatic or manual and each strategy has pros and cons. Automated tools do not enforces all scenarios; some will require manual assessment [14].

Penetration Testing – This is a set of techniques and tools used together to test the software application dynamically against design flaws or vulnerabilities. This activity is important to guarantee that the software or its infrastructure do not have any potential problem that can be exploited in a particular way and change its behavior on the fly.

Security Operations – It is very important to log the user activity into software system usage. Even more important is to maintain that data in a correct and protected manner, to ensure that the attacker or attack activities can be tracked down after the attack attempt.

IV. Software Assessed

The assessed medical device is responsible to monitor vital signs from a patient and send collected information to an Android smartphone. This system, showed as a diagram at Figure 4, is divided into a Body Sensor Network (Figure 5a), composed by a sensor set that monitor vital signs, a Coordinator sensor that collects information from body sensors in a regular basis and re-send that data to the smartphone and the Monitor software (Figure 5b) that evaluate patient conditions time to time. For this software /

equipment no injury is possible, arranging it into Class A classification, according to [2].

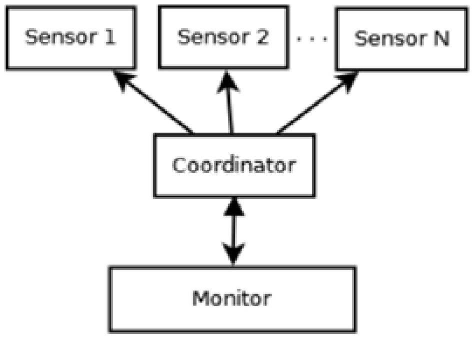


Figure 4. Monitor System Diagram.

This medical device is developed as a research project of the Software Engineer Group from Computer Science Department at University of Brasilia and was provided as a courtesy for this assessment. The research group responsible for developing the monitor system is not the same group that performed the software analysis. Notice that the only part assessed in this work is the Monitor software. Mechanical parts, sensors and smartphone hardware are not part of that analysis. Monitor software was developed in Java Language to run in Android devices.

This software uses the *Software Product Line (SPL)* methodology to build reusable components. In *SPL*, each product is a different piece of software that has some common artifacts in its structure [17]. In medical area, the use of *SPL* methodology brings some problematic issues related to validation and verification of safety characteristics. So, the research team [17] built Monitor software to verify the use of a parametric validation checking model to ensure safety properties (availability, reliability, security, integrity and maintainability). It was done because all medical device software must have dependable and reliable characteristics to guarantee safety.

This device and its related software were a good candidate to security evaluation since the software was in early development state and uses an unusual development methodology for medical devices. It is especially interesting to see if security activities really fit into a new development methodology or perspective.

V. Analysis of the Software Based on the Methodology Proposed

The analyzed software was not plan or built with any security touch point in mind. To help improve the software security and safety was performed an evaluation to propose and add touch point activities into software lifecycle, especially into building steps. Those touch points could be added into software lifecycle at any time, but it is better to do it when the software contains those activities from the scratch.

There are some steps to assessment take place. These steps can be related with one or more touch points each time and they were performed to track the assessed software into a security lifecycle.

Just for the record, safety practices listed at ISO 62304:2006 and other standards will not be ignored here but overlapped by security perspectives. It will be added at software process to increase safety and establish security. For example, risk analysis, abuse cases and risk-based tests are already present in safety related processes and this work will bring security concerns to these activities.

Abuse cases are related with vulnerabilities and flaws. For this analysis were defined SQL injection vulnerability and authentication and authorization problems as abuse cases. SQL injection, for example, could reveal validation problems in application. That is a common vulnerability in software [10, 14, 15] and must be mitigated. Authentication and authorization problems could show problems related to software design flaws [15].

Risks, in a security perspective, are directly related to software failures or vulnerabilities. The risk for SQL injection vulnerability is information disclosure and for authentication and authorization problems are non-legitimate user accessing and exploring the application. The risk-based security tests will be related to the risks specified. In code phase, these risks must be mitigated to ensure no path for exploitation.

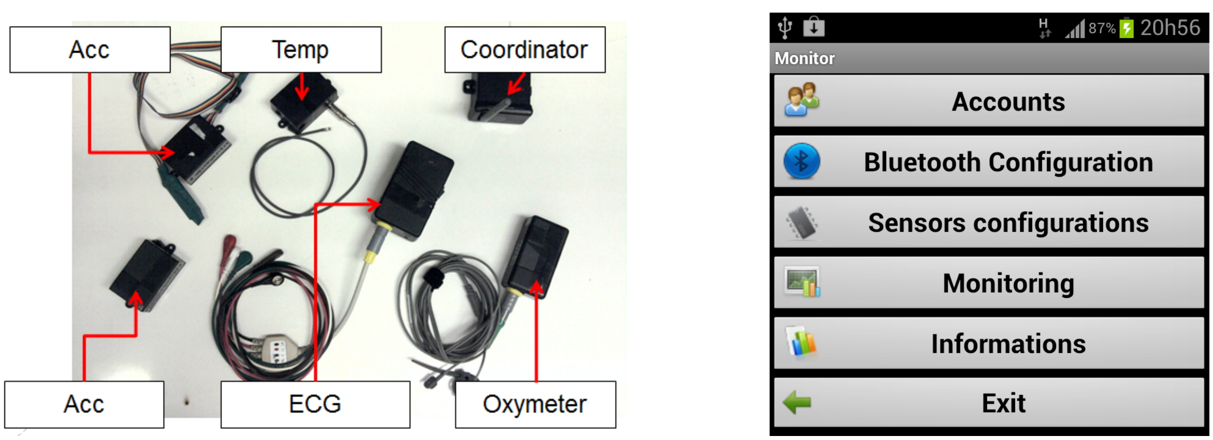


Figure 5. (a) Body Sensor Network (*left*). (b) Monitor software interface (*right*).

Code review is an important control strategy. This methodology comprises, at least, the following elements: Track user-controllable entry point data and review source code responsible for process it, search evidences to ensure that there is no vulnerability related to risks in source code and look for known patterns for common vulnerabilities and perform a line-by-line review of risky code to understand application logic and flaws that may exist [14, 16].

The code review phase could use tools, but it is necessary keep in mind that tools does not do all work. Manual review is always required.

Problems related with abuse cases and with risks specified above were found in Monitor software source code when performed a detailed code review. Field validation and authentication controls are not properly implemented. Examples of vulnerabilities found in source code review are shown in Table 1.

TABLE I. SOME PROBLEMS FOUND IN CODE REVIEW

#	Vulnerabilities		
	Flaws	***Class Name***	***Line Number***
1	Logging of user activity	Global (Many Classes)	N/A
2	No validation on input field	AccountMaintainActivity.java	142
3	Persistent Command Injection	UserDAO.java	119

To confirm that problems found in source code review could really be exploited, a penetration test must be performed. There are, at least, three phases involved in penetration testing: test preparation, test and test analysis as shown in Figure 6.

First phase is related to scope, objectives, timing and duration of the test. All legal agreements must be arranged during this phase. Second phase is considered the bulk of penetration test process. This phase involves application information gathering, vulnerability analysis and exploits. Results are investigated and analyzed in the last phase. The final report generated must be comprehensive and systematic [18].

Security operation is concerned with platform problems that could happen while software is working. Monitor software must be configured following Android Platform security specifications and requirements, as show in [13]. Examples of described requirements are data protection, cryptographic practicalities and use of protected communication channels.

This work is not confirming source code review with a penetration test since application is on early development stage. As soon as Monitor software development starts follow a security development plan, regular dynamic evaluation will be performed as soon as software becomes mature.

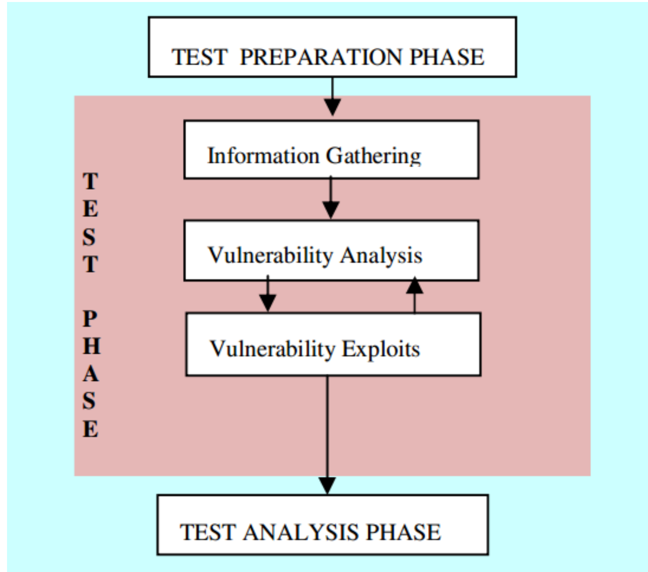


Figure 6. Penetration Test Phases [18].

Despite code review was not confirmed with penetration test, the common flaws shown in Table 1 are enough to demonstrate that touch points must be considered in software development lifecycle. A hacker or an attacker with moderated knowledge can exploit these software flaws easily.

VI. CONCLUSION AND FUTURE WORK

This assessment showed the importance of observing the security aspects in the software development lifecycle. The standards used for regulation of medical device software do not take into account security concerns. These aspects can make all difference in final software security and also in patient safety.

It is responsibility of *QSRs* deal with security concerns clearly. In general, standards for normalization of validations and verification are worried about functional aspects of software operation. Security issues are generally collateral problems that persist in all phases of software lifecycle, until software finishes its production life.

The monitor software used in the analysis was not designed, and as consequence, built with security concerns. So, every kind of security issue can appear in assessment. Since assessed software is in earlier stage of development, it is easier to map problems, flaws, issues and vulnerabilities and create a plan to mitigate them.

Generally, this kind of assessment produces lots of confidential results, and it is difficult to show them without brake non-disclosure agreements and/or reveal sensitive information about software internal structure. More relevant results were discussed directly with design and implementations teams involved in research project.

Unfortunately, securities problems are only solved when entire team involved in software construction are conscious about how it can affect in software operation.

To create this kind of conscience lots of actions are important. But, only organizations that have a security culture and security personal with secure coding and assessment skills can address these actions correctly.

In next steps, a complete penetration test will be performed, trying to exploit vulnerabilities found in code reviews and confirming that risks mapped were mitigated.

ACKNOWLEDGMENT

The authors would like to thank Software Engineer Group from Computer Science Department at University of Brasilia for allowing the analysis performed at Monitor software. This software is part of a research project entitled *Ambient Assisted Living Product Lines.*

REFERENCES

[1] H. Fraser, Y. Kwon, and M. Neuer, The future of connected health devices, IBM Institute for Business Value, New York, 2011.

[2] IEC 62304, Medical device software – Software life cycle processes, 1st ed, Geneva, 2006.

[3] ISO 27799, Health informatics – Information security management in health using ISO/IEC 27002, 1st ed, Geneva, 2008.

[4] J. Radcliffe, "Hacking Medical Devices for Fun and Insulin: Breaking the Human SCADA System", Black Hat Conference, Las Vegas, 2011, http://media.blackhat.com/bh-us-11/Radcliffe/BH_US_11_Radcliffe_Hacking_Medical_Devices_WP.pdf, 01.01.2013

[5] D. Halperin, et al., "Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses", Proceedings of the IEEE Symposium on Security and Privacy, Oakland, 2008, pp. 129-142, doi:10.1109/SP.2008.31.

[6] S. R. Rakitin, "Coping with Defective Software in Medical Devices", Computer Magazine - IEEE Computer Society, v. 39, 2006, n. 4, pp. 40-45, doi: 10.1109/MC.2006.123.

[7] D. A. Vogel, Medical Device Software Verification, Validation, and Compliance, Boston: Artech House, 2011.

[8] G. McGraw, Software security: building security in, Boston: Addison Wesley Professional, 2006.

[9] U.S. Food and Drug Administration, Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices. 1st ed, New Hampshire, FDA, 2005.

[10] Open Web Application Security Project, OWASP Top Ten – 2010 The Ten Most Critical Web Application Security Risks, CC:OWASP, 2010.

[11] K. Hall, "Developing Medical Device Software to IEC 62304", European Medical Device Technology Magazine, v. 1, n. 6, June 2010.

[12] R. J. Anderson, Security Engineering: A Guide to Building Dependable Distributed Systems, 2nd ed, Indianapolis: Wiley Publishing inc, 2008.

[13] J. Six, Application Security for the Android Platform, 1st ed California: O'Reilly Media, Inc, 2012.

[14] F. Long, D. Mohindra, R. C. Seacord, D. F. Sutherland, and D. Svoboda, The CERT Oracle Secure Coding Standard for Java, 1st ed, Michigan: Pearson Education Inc, 2012.

[15] D. Stuttard and M. Pinto, The Web Application Hacker's Handbook: Finding and Exploiting Security Flaws. 2nd ed Indianapolis: Wiley Publishing inc, 2011.

[16] M. Paul, Official (ISC)2 Guide to the CSSLP, 1st ed, Florida: CRC Press, 2011.

[17] V. Nunes, P. Fernandes, V. Alves, and G. Rodrigues, "Variability Management of Reliability Models in Software Product Lines: an Expressiveness and Scalability Analysis", I: SBCARS - Simpósio Brasileiro de Componentes, Arquitetura e Reutilização de Software, Natal - Brazil, 2012, pp. 113 - 122.

[18] A. G. Bacudio, X. Yuan, B. B. Chu, and M. Jones , "An Overview of Penetration Testing", International Journal of Network Security & Its Applications (IJNSA), v.3, 2011, n.6, pp. 19-38, doi: 10.5121/ijnsa.2011.3602.

"Handreha": A new Hand and Wrist Haptic Device for Hemiplegic Children

Mohamed Bouri, Charles Baur, Reymond Clavel

Ecole Polytechnique Fédérale de Lausanne
Laboratoire de Systèmes Robotiques (LSRO)
CH 1015, Switzerland
Mohamed.Bouri@epfl.ch, Charles.baur@epfl.ch ,
Reymond.clavel@epfl.ch

Milan Zedka, Christopher John Newman

Nestlé Hospital CHUV
Service of Neurology and Pediatric Neurorehabilitation
CH 1011 Lausanne, Switzerland
Christopher.Newman@chuv.ch, Milan.zedka@chuv.ch

***Abstract*— This paper presents the development of a new haptic device for hand rehabilitation for hemiplegic children. "Hanreha" has been developed at the EPFL thanks to the interest of the "Neurology and Pediatric NeuroRehabilitation" Service of the CHUV (Centre Hospitalier Universitaire du Canton de Vaud). The novelty of this device is that it is a 3 degrees of freedom desktop system, supporting pronation/supination, flexion/extension and grasping hand movements and it is totally dedicated, in its current state, to children. The kinematics and the construction aspects of this desktop device are presented. Its different advantages are discussed to point out the benefits of this structure. Control and force feedback aspects combined with virtual reality are also presented. A prototype of the "Handreha" is realized and presented and the performances discussed. The first evaluations with hemiplegic children really show that the mechanical design of the device fits the targeted specifications**

***Keywords*- Hand Rehabilitation; wrist; hemiplegic; children; force feedback; control; virtual reality.**

I. INTRODUCTION

Rehabilitation therapy for paraplegics, quadriplegics and hemiplegics often consists of mobilization exercises of the affected limb by a therapist (Figure 1). The exercises may have different objectives depending on the pathology. They may seek to reduce hypertonia, increase the joint range of motion, improve the plasticity behavior of the limb, increase muscular strengths, decrease spasms or attain various other objectives [1] [2]. This conventional mode of therapy gives good results, but it is severely limited by the fact that a therapist is needed for each single patient during the total period of exercise. Increasingly, the cost of this to the health care system is enormous, and the need for therapy far exceeds the availability of trained therapists.

Currently available rehabilitation devices are robot-like structures [3][4][5][6][7]. Two examples are illustrated in Figure 2 for both lower limbs (the MotionMaker) and upper limbs (Armeo).

Figure 1. Therapist moving the fingers of a hemiplegic child (Kawahira method [8])

The robotized devices provide repetitive, precise and totally instrumented mobilization of the limbs.

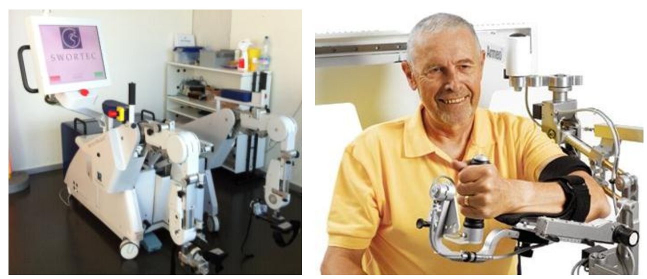


Figure 2. Examples of Rehabilitation robots: (Left) The MotionMaker for lower limbs [9], (Right) The ARMEO for upper limbs [5].

Figure 3. **Handreha**: The developed desktop hand rehabilitation device

In this paper, we are interested in the task of hand and wrist rehabilitation and we will present the developments of a 3 degrees of freedom (DOF) robot: the "Handreha", developed at the EPFL (Figure 3) thanks to the interest and at the request of the "Neurology and Pediatric Neuro-Rehabilitation" Service of the CHUV hospital. The novelty of this device is that it is desktop system, supporting pronation/supination, flexion/extension and grasping hand movements and it is totally dedicated, in its current state, to children.

"Handreha" has been developed for hand rehabilitation of hemiplegic children from 7 to 14 years old. The other aspect of this work presented in the section IV-B is the development of a virtual labyrinth game with haptic force feedback. The objectives of the game are motivating the children and keeping them using the rehabilitation exercises.

II. Overview of some hand rehabilitation devices

Haptic interfaces allowing force feedback are required in order to improve interaction through a virtual environment. Several devices have been developed. However, these interfaces are mostly focused on the rehabilitation of the shoulder and the arms, but not the hand. There are more hand rehabilitation systems based on "gloves" [11] [12] [13]; but, unfortunately, even if they are more accurate, they are more complex and expensive (Figure 4-Right).

The HWARD (Figure 4-Left) focuses on the rehabilitation of the hand only. It is a robot with 3 DOF for the movement of the fingers, thumb and flexion / extension of the wrist. It can assist the patient in grasping real objects without any virtual reality support. This is not a haptic system but one of the only hand rehabilitation robots developed until now [14].

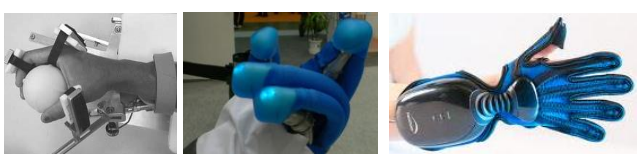


Figure 4. Hand Rehabilitation: (Left) the HWARD, (Mid) The hand glove from Panasonic, (Right) Hand glove HandTutor [13]

The Gentle/G (Figure 5) allows the rehabilitation of the arm, shoulder and hand. The movements of reaching and grasping are exercised through a 6 active DOF and 3 passive DOF: 3 DOF for positioning the active arm, 3DOF passive positioning of the hand and 3 DOF for the assets reaching movement. This robot anthropometric settings [16]. Other similar systems are the "Robotherapist" presented in [17] and [18] and the Armin [5] (Figure 2) commercialized under the name of ARMEO by the company Hocoma (CH).

Figure 5. The Gentle/G system during a rehabilitation exercise

As we previously noticed, other different devices exist and are more dedicated to the rehabilitation of the entire upper extremities.

Here are some examples.

- The MIME device that is built using a PUMA robot for a 6 DOF hand rehabilitation [19].
- The MIT-Manus [20] which is now commercialized by the company Inmotion. A new variant provides the flexion/extension and pronation/supination movements and also the radial and ulnar deviations of the wrist.
- The HENRIE uses the robot Haptic MASTER (3 DOF) for arm rehabilitation. A supplementary passive grasping DOF (spring based) is added. Thus, the stiffness of the "virtual object" felt is constant. With the Haptic MASTER, the weight of the object is felt on the wrist [21].
- The SEAT (Figure 6) for arm rehabilitation uses a steering wheel for the measurement of the applied force applied by each arm (left and right). It is equipped with a motor that assists the plegic member or makes it work longer due to the application of a resistant torque [22].

Figure 6. The SEAT System

- Another interesting device (Figure 7) is the ReHapticKnob developed by the RELAB [26]. It is a 2 degrees of freedom desktop and compact device that may be used for the assessment and therapy of grasping. It provides both pronation/supination and grasping movements.

Figure 7. The ReHapticKnob

III. THE MECHANICAL DESIGN OF THE "HANDREHA"

A. *Objectives*

The principal objectives of our development have been to develop a light desktop, three DOF hand rehabilitation device for hemiplegic children. Each degree of freedom is actuated and the device is able to be used for haptic purposes and interfaced with virtual reality environments.
The 3 supported degrees of freedom are as follows (Figure 8):

- The rotation of the forearm (Rx) over 180°, namely the pronation / supination.
- The orientation of the hand (Rz) over 180°, namely the flexion / extension of the wrist.
- The grasping motion (Θ) over 60° that allows opening and closing of the gripper.

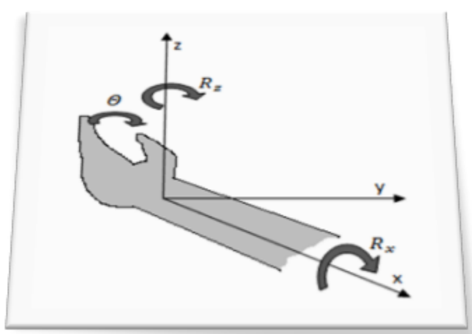


Figure 8. Required degrees of freedom of the HANDREHA device

This paragraph presents the haptic interface Handreha device developed by EPFL Its development has been motivated by the availability of a cost-affordable desktop device designed for hemiplegic children. Both "**rehabilitation**" and **"assessment"** purposes have been targeted. The progress in wrist capabilities by increasing the motion, force and coordination are the principal objectives of "HANDREHA".
Handreha can be positioned on a table and may be configured for the right hand as well as for the left hand.
All the joints are motorized and instrumented with position incremental encoder sensors to allow controlled force/position feedback.

B. *Anthropomorphic requirements*

Prado-Leon et al. [24] surveyed anthropometric data of Mexican children. Linghua et al. [25] also measured anthropometric data of Chinese children. Table III and Table IV, respectively, summarize this data in the Appendix A.

Nevertheless, we also carried out a series of anthropomorphic measurements conforming to our required objectives which concern the tasks of flexion/extension and grasping (Figure 9). These measurements are compiled in the (Table I). 9 children were implicated in this series of measurement to comfort the literature data.

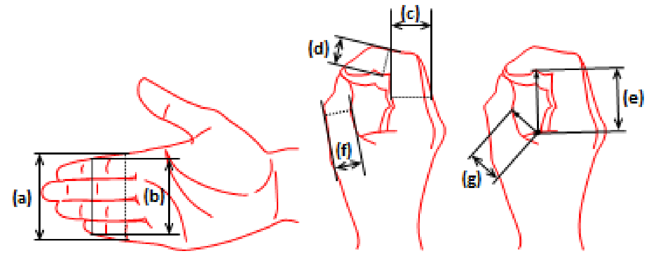


Figure 9. Anthropomorphic measured parameters

TABLE I. ANTHROPOMETRIC MEASURED HAND PARAMETERS

Dimensions (mm)		Subject	A	B	C
		Age (y/o)	6	7	9
Finger	Breadth at proximal	**(a)**	51	53	58
	Breadth at middle	**(b)**	47	51	52
	Thickness at proximal	**(c)**	19	21	21
	Thickness at middle	**(d)**	15	15	15
	Length to DIP from MP	**(e)**	33	38	39
Thumb	Thickness	**(f)**	15	14	16
	Length to IP from MP	**(g)**	18	20	19

To develop "Handrea", our desktop device for hemiplegic children from 6-14 yo, we decided to adopt the following parameters (Table II). For h and i parameters see Figure 10.

TABLE II. ANTHROPOMORPHIC HAND REQUIREMENTS

Dimensions (mm)			
Finger	Breadth at proximal	**(a)**	50-75
	Breadth at middle	**(b)**	45-70
	Thickness at proximal	**(c)**	18-25
	Thickness at middle	**(d)**	15-20
	Length to DIP from MP	**(e)**	30-45
Thumb	Thickness	**(f)**	15-25
	Length to IP from MP	**(g)**	18-30
Elbow	Elbow to wrist	**(h)**	150-300
	Elbow to ground	**(i)**	450-600

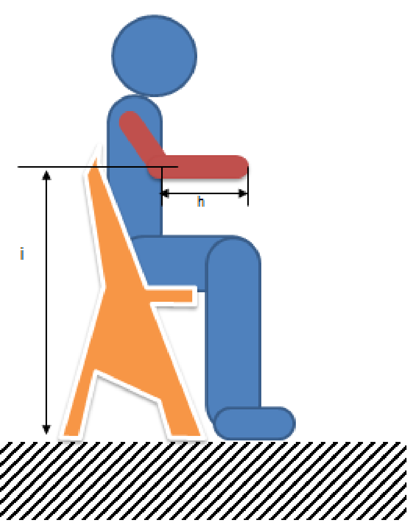


Figure 10. Seating configuration and parameters (Table II)

 ISBN: 978-1-61208-250-9

C. *Ergonomic requirements*

The grasping mechanism has been chosen to be closest to scissors principle (Figure 11).

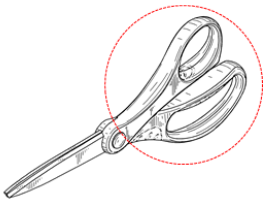
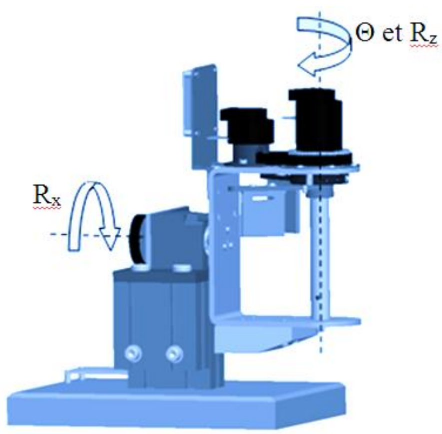


Figure 11. Scissors handling

Scissors are often used in occupational therapy (OT) and they provide more natural grasping movements than the use of simple plate type grasping.

D. *Mechanical design*

Handreha has been designed as a serial kinematics with one motor on the frame ensuring the pronation/supination degree of freedom (Rx). The other motors are on the mobile part and their movements are combined to carry out the flexion/extension and grasping (Figure 12).



6

Figure 12. The mechanical structure

In terms of torques, the joints respectively allow the permanent torques of $\boldsymbol{M_{xrms}}$ **= 0.2Nm** and $\boldsymbol{M_{zrms}}$= $\boldsymbol{M_{grasp\_rms}}$= ***0.3Nm*** (which corresponds to a grasping force of **4N** or 400 gram-forces).

Children with Hemiplegia may have muscle spasms and this leads to the possibility of the presence of radial and axial transient forces of up to 25 [N].
There is no need for fast motions. The maximal required velocity is around 10 rpm (180° per 3 seconds). The maximal acceleration is 20 rpm/s/s (assuming that we reach the maximal velocity in 0.5s).

This structure can be divided into two parts. The first is the rotation of the forearm and the second is the grabbing system composed of the two other movements. As we can observe, the grasping motion is coupled to the rotation of the wrist around the z-axis.

1) *The rotation of the forearm*

The 180° rotation of the forearm is simply realized by one active pivot driven by a Maxon EC45 motor with a belt transmission assuring a convenient gearing to make available the required torque.

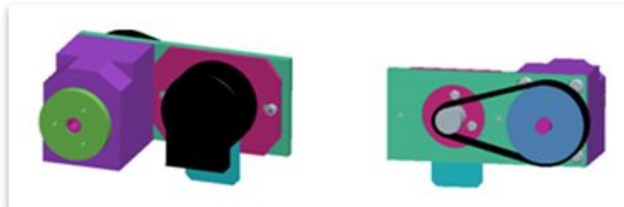


Figure 13. The rotation of the forearm

2) *The grabbing system*

The grabbing system (hand orientation and the grabbing movement), is composed of two active rotations. The first (the hand orientation Rz) is driven by motor number 2 (Figure 14 and Figure 15). This motor drives the stator of motor number 3 through a belt. In order to transmit this rotation to the hand of the hemiplegic child, two arms forming a gripper are fixed to the stator of motor 3 (stator arm) and also to the rotor of motor 3 (rotor arm). Therefore, when the hand of the child turns with a fixed angle of aperture, both arms move simultaneously.

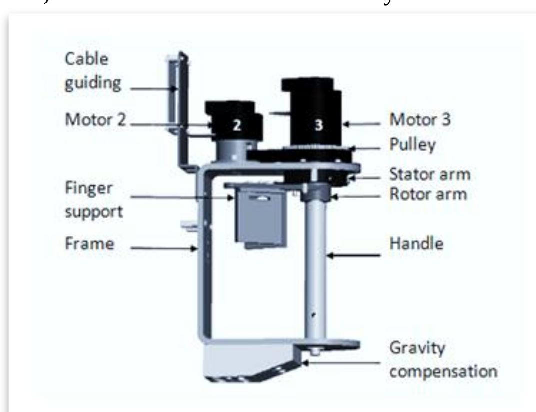


Figure 14. The grabbing system details with the previous version of Handreha

As for the grabbing movement, it is driven by motor number 3 through planetary gearing. Thus, when the child closes or opens its hand, only the rotor of motor number 2 moves.

Figure 15. The grabbing system

 ISBN: 978-1-61208-250-9

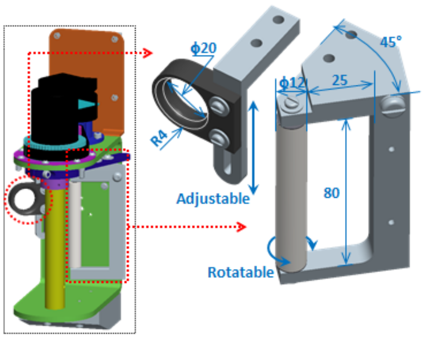


Figure 16. The new grabbing system of Handreha **conforming to the scissors principle**

IV. REALIZATION AND CONTROL OF HANDREHA

A. The prototype

The photos below show the HANDREHA prototype. This prototype is intended to be fixed on a table to allow mobilization through a force feedback emulation using a virtual reality application and different interaction models.

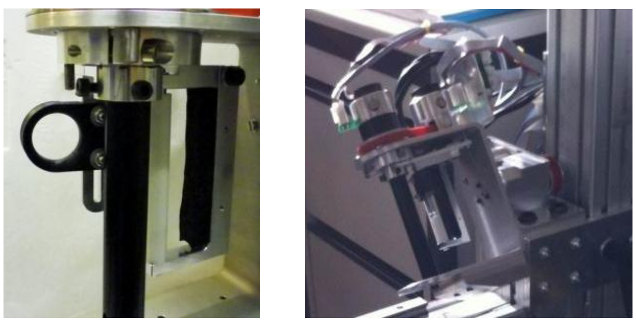


Figure 17. The prototype of HANDREHA

The device has been mounted on a mobile frame to allow its use for clinical trials. This also has the advantage of having the electronics on the frame.

This frame allows the adjustment of the elbow rest to conform to the seat height (Figure 18). Another interesting aspect which has been considered is the securing of the elbow movements by guiding it (the elbow) through a linear movement in the direction of the pronation axis. This totally removes the elbow compensation movements in flexion/extension of the wrist (Figure 20).

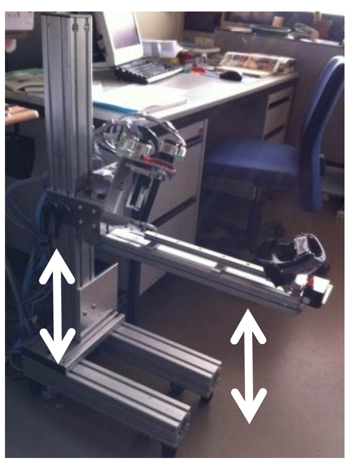


Figure 18. Mobile frame and height adjustment

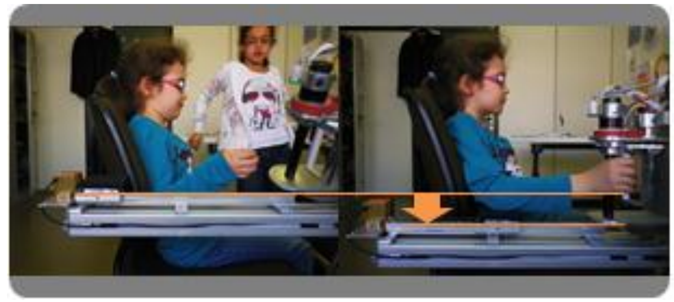


Figure 19. Elbow rest Height adjustment with a 7 years girl

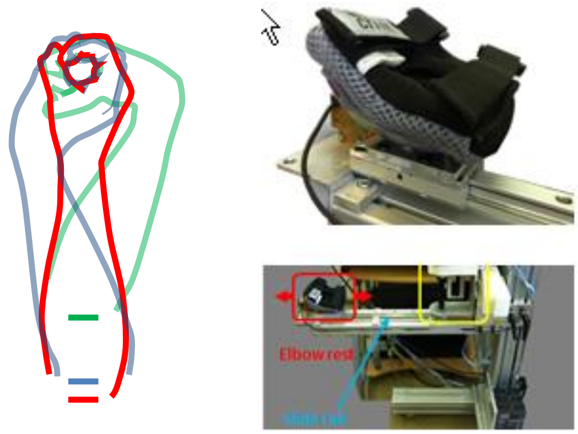


Figure 20. Elbow rest guiding to avoid compensating movements in flexion

B. The controller and the virtual application

The controller has to impose the values of applied torques by each motor at each joint and ensures a desired behavior (stiffness, linear or nonlinear viscosity, …) with respect to a chosen environment.

This is carried out both by the servo-amplifier of the motors that are configured in torque mode (the torque mode bandwidth is 1 KHz) and an additional virtual application. The virtual application sets up the desired force behavior which is function of the measured joint positions and velocities each 1 ms (Figure 21). More details concerning this PC based open controller may be found in [26].

To be effectively adapted to manipulation by children and motivate them to use the haptic device, a dedicated application has been developed. This application is based on a virtual labyrinth with different geometric configurations.

This application is associated with force feedback implemented both in the grasping operation (by implementing different grasping viscoelastic effects) and when moving inside the labyrinth (by implementing different viscoelastic effects during the motion in the labyrinth and by simulating the stiff contacts with the walls).

 ISBN: 978-1-61208-250-9

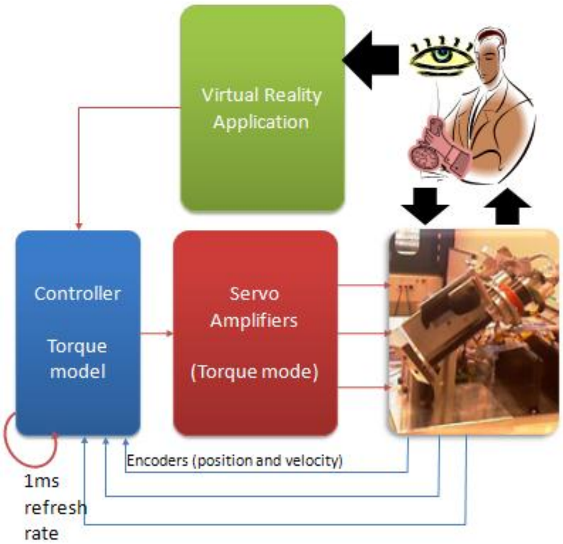


Figure 21. The haptic force control concept

The child has to grasp a ball and manipulate it through the paths of the labyrinth to reach a target allowing the child to increase its score. The configurations correspond to different levels of difficulty (Figure 23), allowing different corresponding wrist angular positions.

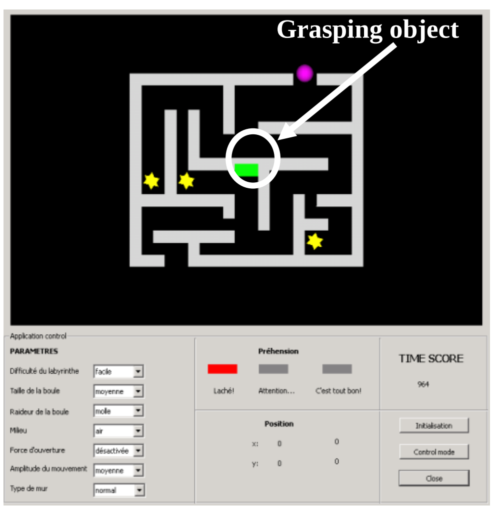


Figure 22. The Virtual Labyrinth

To manage the grasping effect, the child has to decrease the diameter of the ball by tightening (more or less) the gripper to pass through the obstacles.

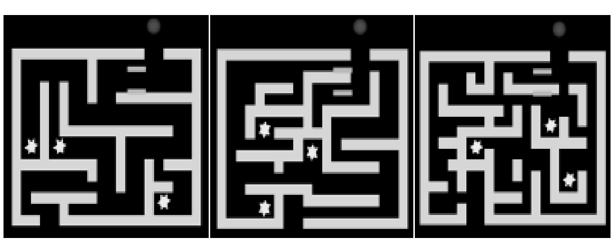


Figure 23. Different labyrinth difficulties

Figure 24 shows a healthy child manipulating the HANDREHA by playing the “Labyrinth game”.

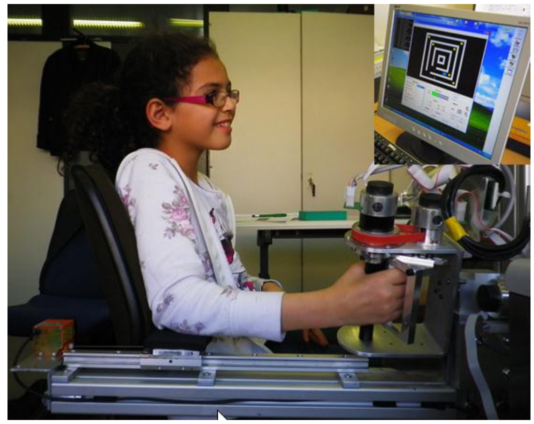


Figure 24. A 9 years sound children manipulating the Handreha

The ball is manipulated through the paths of the 2D labyrinth and a coordinate transformation is carried out to transform the rotation movements Rx and Rz into translational movements (respectively forward-straight / back-straight and forward-right / forward-left). Different wall effects are also available to differentiate the sensations and make the game more attractive.

A first clinical trial of the device was carried out with three hemiplegic children (7, 10 and 12 years) at the service of “Unité de Neurologie et Neuroréhabilitation Pédiatrique, Hôpital Nestlé CHUV, Switzerland” to check the mechanism and the dimension parameters. All the manipulated variables (maximum torques, articular angles, velocities, scores) have been saved in a database in order to carry out a systematic and a quantified progress of the capabilities during the exercises. These tests have concluded to the improvements proposed in the current paper.

V. CONCLUSION

This paper presented the mechanical development of “HANDREHA”, a hand and wrist haptic device for hemiplegic children. This device has been particularly developed as desktop robotic system and provides three degrees of freedom: pronation/supination, flexion/extension and grasping. Force feedback aspects have also been considered to enlarge the implementation of different rehabilitation strategies.

HANDREHA is an academic prototype developed by the Laboratory of Robotic Systems (EPFL). It is an affordable device and very flexible in terms of adaptation to desktop applications and for hospital clinical therapeutic evaluations. It is adapted for children from 7 to 14 years old and allows elbow rest height adjustment and configuration of hand parameters. The elbow compensation movement is totally removed trough guided elbow movement.

It is now at the hospital CHUV of the Canton of Vaud in Lausanne and a second series of clinical trials is programmed. These tests will use the developed application for tests with different hemiplegic children.

REFERENCES

[1] A. Joel DeLisa, Physical Medicine and Rehabilitation: Principles and Practice, 4th edition, 2100 pages, 2004.

[2] H. Janet Carr, and B. Roberta, Neurological Rehabilitation: Optimizing Motor, 364 pages, 2003.

[3] J. Stein, R. Harvey, R. Macko, C. Winstein, and R. Zorowitz, Stroke Recovery and Rehabilitation, Ed Demos MEDICAL, 797 pages, 2009.

[4] P. Métrailler, V. Blanchard, I. Perrin, R. Brodard, R. Frischknecht, C. Schmitt, J. Fournier, M. Bouri and R. Clavel, Improvement of rehabilitation possibilities with the MotionMaker™, The first IEEE / RAS-EMBS International Conference on Biomedical Robotics and Mechatronics, BioRob, Pisa Italy, 2006,

[5] T. Nef, M. Mihelj, and R. Riener, ARMin: A robot for patient-cooperative arm therapy, Med. Biol. Eng. Comput., Vol. 45, pp. 887–900, 2007.

[6] C. Schmitt, P. Métrailler, A. Al-Khodairy, R. Brodard, J. Fournier, M. Bouri, and R. Clavel, The Motion Maker™: a Rehabilitation System Combining an Orthosis with Closed-Loop Electrical Muscle Stimulation, 8th International Workshop on Functional Electrical Stimulation, pp. 117-120, Vienna Austria, 2004.

[7] G. Charles Burgar, S. Peter Lum, C. Peggy Shor, and H.F. Machiel Van der Loos, Development of robots for rehabilitation therapy: The Palo Alto VA/Stanford experience, Journal of Rehabilitation Research and Development,Vol. 37 No. 6, pp. 663-673, 2000.

[8] K. Kawahira, M. Shimodozono, A. Ogata, and N. Tanaka, Addition Of Intensive Repetition Of Facilitation Exercise To Multidisciplinary Rehabilitation Promotes Motor Functional Recovery Of The Hemiplegic Lower Limb, J Rehabil Med; Vol 36, pp. 159–164, 2004.

[9] Y. Stauffer, Y. Allemand, M. Bouri, J. Fournier, R. Clavel, P. Métrailler, and R. Brodard, Clinical trials with the WalkTrainer: Preliminary results, Journal of Biomechanics, Volume 41, pp. S136-S136, 2008.

[10] Y. Stauffer, M. Bouri, C. Schmitt, Y. Allemand, S. Gnemmi, J. Fournier, R. Clavel, P. Métrailler, and R. Brodard, Cyberthèses, Mise en œuvre d'un nouveau concept de rééducation pour paraplégiques et hémiplégiques, Journal Européen des systèmes automatisés, vol. 41, num. 2, pp. 261-278, 2007.

[11] P. C. Kung, C. K. Lin, M. S. Ju, Neuro-rehabilitation robot-assisted assessments of synergy patterns of forearm, elbow and shoulder joints in chronic stroke patients, Clinical Biomechanics, Volume 25, Issue 7, pp. 647-654, 2010.

[12] T. Markow, T., Ramakrishnan, N., Huang, K., Starner, T., Eicholtz, M.; Garrett, S., Profita, H., Scarlata, A., Schooler, C., Tarun, A., Backus, D., Mobile Music Touch: Vibration stimulus in hand rehabilitation,Pervasive Computing Technologies for Healthcare (PervasiveHealth), 4th International Conference, Munich, 2010.

[13] E. Carmeli, J. J. Vatine, S. Peleg, G. Bartur, and E. Elbo, Upper limb rehabilitation using augmented feedback: Impairment focused augmented feedback with HandTutor, Virtual Rehabilitation International Conference, pp 2200-2202, Israel, 2009.

[14] C. D. Takahashi, L. Der-Yeghiaian, A robotic Device for hand motor therapy after stroke. IEEE 9th international conference on rehabilitation robotics, pp. 17-20, USA, 2005.

[15] L. A Malone, P. L Gervaisb, R. S Burnhamc, M. Chanc, L. Millerd and R. D. Steadwardb, An assessment of wrist splint and glove use on wheeling kinematics, Clinical Biomechanic, Volume 13, Issue 3, pp. 234-236, 1998.

[16] R. C. V. Loureiro, and W. S. Harwin, Reach and grasp therapy: Design and control of a 9-DOF robotic neuro-rehabilitation system, IEEE 10th international conference on rehabilitation robotics, pp. 757-763, Netherlands, 2007.

[17] K. Oda, T. Kikuchi, Y. Ohyama, S. Isozumi, K. Tamida, and J. Urusho, Improvement of Mechanism of Upper Extremity Exercise Support System "Robotherapist" and Development of Application Software, IEEE International Conference on Robotics and Biomimetics (Robio), pp. 2515-2520, China, 2009.

[18] J. Furusho, N. Shichi, X. Hu, T. Kikuchi, K. Nakayama, C. Li, Y. Yamaguchi, A. Inoue and U. Ryu, Development of a 6-DOF Force Display System with High Safety and its Application to Rehabilitation, Proceedings of the 2006 IEEE International Conference on Mechatronics and Automation, pp. 962-967, China, 2006.

[19] P. S. Lum, C. G. Burgar, P. C. Shor, M. Majmundar, M. Van Der Loos, Robot-Assisted Movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. Arch Phys Med Rehabil; Vol 83, pp. 952-957 , 2002.

[20] H. I. Krebs, B. T. Volpe, D. Williams, J. Celestino, S. K. Charles, D. Lynch, and N. Hogan, Robot-Aided Neurorehabilitation : A Robot For Wrist Rehabilitation, IEEE Trans. On Neural Systems And Rehabilitation Engineering, Vol. 15, No. 3, pp. 327-335, September 2007.

[21] M. Mihelj, J. Podobnik, and M. Munih,. HEnRiE – Haptic Environment for reaching and grasping exercise. 2nd IEEE International conference on biomedical robotics and biomechatronics, pp. 907-912, USA, 2008.

[22] M. J. Johnson, H. F. M. Van der Loos, C. G. Burgar, and L. J. Leifer, Driver's SEAT : Simulation environment for arm therapy. ICORR international conference on rehabilitation robotics, Stanford, USA, 1999.

[23] J. C. Metzger, O. Lambercy, D. Chapuis, and R. Gassert, Design and Characterization of the ReHapticKnob, a Robot for Assessment and Therapy of Hand Function, in Intelligent Robots and Systems Conf, pp. 3074-3080, San Francisco, CA, USA, 2011..

[24] L. Prado-Leon, R. Avila-Chaurand, and E. Gonzalez-Munoz, Anthropometric study of Mexican primary school children, Applied Ergonomics, Vol. 32, pp. 339-345, 2001.

[25] L. Ran, X. Zhang, C. Chao, T. Liu, and T. Dong, Anthropometric Measurement of the Hands of Chinese Children, Digital Human Modeling, HCII2009, LNCS5620, pp. 46-54, 2009.

[26] M. Bouri, and R. Clavel, A Windows PC based robot controller: An open architecture. International Symposium on Robotics, Tokyo, Japan, 2005.

VI. APPENDIX

A. Hand Anthropomorphic data (in mm)

TABLE III. ANTHROPOMETRIC HAND DATA OF MEXICAN CHILDREN [24]

Dimensions (mm) — Age	6		7		8		9		10		11	
Sex	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Hand length	130	129	135	134	141	139	146	146	151	153	158	160
Hand palm length	73	73	77	76	80	78	83	82	86	86	90	90
Hand breadth at metacarpal	60	58	62	60	64	62	66	65	68	67	72	71
Grip diameter	26	27	28	28	29	29	30	31	31	32	33	34

TABLE IV. ANTHROPOMETRIC HAND DATA OF CHINESE CHILDREN [25]

Dimensions (mm)	4-6 y/o		7-10 y/o		11-12 y/o	
	Boys	Girls	Boys	Girls	Boys	Girls
Hand length	124.1	122.0	144.2	142.7	161.0	161.3
Hand breadth at metacarpal	58.4	56.5	65.5	63.4	71.8	70.0
Palm length	71.0	69.3	82.3	80.7	91.8	90.9
Index finger length	48.2	48.0	56.0	56.2	62.3	63.5
Thumb length	39.2	38.5	45.9	45.9	51.7	52.3
Middle finger length	53.8	53.6	62.4	62.6	69.6	70.9

Fundamental Study to Consider for Evaluation of A Welfare Device

Hiroaki Inoue/Tokyo University of Science, Suwa
Research course of Engineering/Management
Chino-city, Japan
jgh12701@ed.suwa.tus.ac.jp

Shunji Shimizu/Tokyo University of Science, Suwa
Department of Electric Systems Engineering
Chino-city, Japan
shun@rs.suwa.tus.ac.jp

Noboru Takahashi/Tokyo University of Science,Suwa
Research course of Engineering/Management
Chino-city, Japan
srl@ed.suwa.tus.ac.jp

Hiroyuki Nara/Hokkaido University
Graduate School of Information Science and Technology
Sapporo-city, Japan
nara@ssc.ssi.ist.hokudai.ac.jp

Takeshi Tsuruga/Hokkaido Institute of Technology
Department of Clinical and Rehabilitation Engineering
Sapporo-city, Japan
tsuruga@hit.ac.jp

Fumikazu Miwakeichi/The Institute of Statical Mathematics
Spatial and Time Series Modeling Group
Tachikawa-city, Japan
miwake1@ism.ac.jp

Nobuhide Hirai/Jichi Medical University
Department of Psychiatry
Shimotsukeshi-city, Japan
nobu@nobu.com

Senichiro Kikuchi/Jichi Medical University
Department of Psychiatry
Shimotsukeshi-city, Japan
skikuchi@jichi.ac.jp

Eiju Watanabe/Jichi Medical University
Department of Neurosurgery
Shimotsukeshi-city, Japan
eiju-ind@umin.ac.jp

Satoshi Kato/Jichi Medical University
Department of Psychiatry
Shimotsukeshi-city, Japan
psykato@jichi.ac.jp

***Abstract*— Recently, Japan (also world-wide countries) has become aged society, and a wide variety welfare device and system have been developed. But evaluation of welfare system and device are limited only stability, intensity and partial operability. So, evaluation of usefulness is insufficient. Evaluation of usefulness is necessity to consider about interaction of human and welfare device. In this paper, we measure load of sitting and standing movement to use EMG (Electoromyogram) and 3D Motion Capture and set a goal to establish objective evaluation method. We think that establishing objective evaluation method is necessity to develop useful welfare device. We examined possibility of assessing load and fatigue from measuring brain activity to use NIRS (Near Infra-Red Spectroscopy). We think that measuring load and fatigue is very important for developing user-friendly welfare device.**

Idea of universal design is widespread in welfare device and system. Measuring require verification of all generations. But, we performed to measure younger subjects as a first step. We think that younger subjects were observed the significant difference, because they had enough physical function. Considering younger subjects as a benchmark is appropriate for creating evaluation method.

Keywords- Evaluation; Movement; Exercise; 3D Motion Capture; NIRS; EMG; Care; Welfare Technology; Usefulwelfare device evaluation; Evaluation method.

I. Introduction

As increasing aging population in Japan and world-wide countries, welfare systems and device are rapidly developing, and various devices are manufactured based on the increased popularity of welfare device and system. Also, the market of welfare device and system are expanding. However, the evaluation method is limited respectively to stability, strength and a part of operability for individual system or device. It means that evaluation methodology for usefulness of them was not established. Therefore, we will attempt to establish the standard to evaluate usefulness for objectively and quantitatively on the basis of cognition such as physical load, reduction of fatigue and postural stability. Especially, in considering universality, it is necessary to measure human movement in daily life. Movement was not measured by using particular device, but routinely-performed movement in daily life.

So, we examined the possibility of evaluation by measuring physical load due to activities of daily living with using 3D Motion Analysis System and EMG. Also, we looked into the possibility of quantitative evaluation of tiredness and load on the basis of brain activity using NIRS. Also, we consider that physical and psychological load are linked to cognition including non-verbal cognition. In this paper, the purpose of experiments are to evaluate motion focusing on sitting and

 ISBN: 978-1-61208-250-9

standing movement, which is usually done in our life by using 3D Motion Analysis System, EMG, NIRS. We consider that human feel physical and psychological load during life motion. We tried to measure physical load by using 3D Motion Analysis System, EMG. Additionally, we tried to measure non-verbal cognition about psychological load by using NIRS.

Subjects were healthy males in twenties, because the elderly people who has various types of disease is inept in quantitative evaluation.

II. Experimental Method

A. Evaluation by using 3D Motion Analysisand EMG

We simultaneously measured 3D position and muscle potential of subject during task by using 3D Motion Analysis System (nac IMAGE TECHNOLOGY Inc. products-MAC3DSYSTEM [1]) and EMG(KISSEI COMTEC Inc. products-MQ16 [2]).

Regarding to measuring 3D position, 8 Infrared cameras were placed around subject, and 27 makers of the body surface were set on the basis of Helen-Hayes Hospital Marker set(Figure 1). In measuring muscle potential, measurement regions were tibialis anterior muscle, gastrocnemius muscle, quadriceps femoris muscle, hamstring, flexor carpi ulnaris muscle, extensoe carpiulnaris muscle, triceps branchii, latissimus dorsi muscle of the right side of the body because these muscle were deeply associated with standing and sitting movement. Also, wireless measurement was used so that subject was constrained as little as possible.As sampling frequency, 3D Motion Analysis System was 100Hz,and EMG was 1kHz.

Subjects were three males aged twenties. They were asked to read and sign an informed consent regarding the experiment.

In this experiment, subject repeated one series of movement, which was to transfer from chair to seat face of welfare device (IDEA LIFE CARE Co. Ltd products-NORISUKEsan [3]) and opposite one with alternating between standing and sitting, at five times per one measurement. Seating face of welfare device, which was designed to assist transfer movement, was manipulated by simple method and appeared on the top of chair.

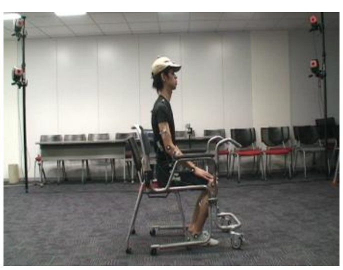


Figure 1. Experimental View of 3D motion Analysis and EMG

Subjects were heard buzzer every one second and kept a constant motion of speed to satisfy certain measuring conditions. Also, they transferred from seat face to chair or conversely every 8 seconds with consideration for movement of elderly persons. Operation of welfare device was performed by the operator other than subject.

B. Evaluation by using NIRS

We measured brain activity during motion with the purpose of establishing evaluation method based on generality (Figure 2).

Subjects were six males aged twenties. They were asked to read and sign an informed consent regarding the experiment. Measurement apparatus was NIRS(SHIMADZU CO. Ltd products-FOIRE3000 [4]). Measurement region was at right and left prefrontal cortex.

1) Measuring brain activity during transfer with standing position(task1)

At this measurement,the subjects used welfare device to perform transferring in a standing position. In this measurement, subject sat on seating face of welfare device appeared on the top of chair after raising hip until kneeling position. Also, subject performed inverse transferring from seating face to chair. Time design was rest(5 seconds), task(10 seconds), and rest(5 seconds). This time design was repeated 30 times. Rest time is to stabilize the brain activity. In the measurement NIRS,

2) Measuring brain activity during transfer with half-crouching position (task2)

At this measurement,the subjects used welfare device to perform transferring in a half-crouch position. In this measurement, the subjects sat on seating face of welfare device appeared on the top of chair after raising hip until kneeling position. Also, the subject performed inverse transfer from seating face to chair.Time design was rest(5 seconds), task(10 seconds) and rest(5 seconds). This time design was repeated 30 times.

In experiments of task1 and task2, operation of welfare device was performed by operator other than subject. Before this measuring, subjects adjusted to transferring by use of welfare device.

3) Measuring brain activity during keeping a half-crouch position(task3).

The subjects performed two tasks at this measurement. During task3-1, subject sat on seating face of welfare device with eyes open. During task3-2, they kept a half-crouch position.

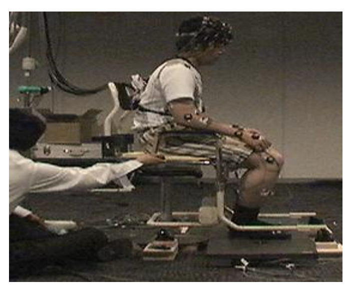


Figure 2. Experimental View of NIRS

Subjects alternated task3-1 and task3-2. Also, subjects took resting time between two types of motion with eyes close. Therefore time design was rest(5 seconds), task3-1(10 seconds), rest(5 seconds), task3-2(10 seconds) and rest(5 seconds). This time design was repeated 15 times.

III. Experimental Result

A. Evaluation by using 3D Motion Analysis and EMG

Figure 3 shows result of transferring which was measured by 3D motion analysis and EMG. In Figure 3, middle trochanter is the height of midpoint between right and left trochanter from the floor. Trunk angle is the forward slope of trunk. Also, following terms are arectifying voltage wave for each eight muscles, which are Tibialis anterior muscle, Astrocnemius muscle, Quadriceps femoris muscle, Hamstring, Triceps brachii muscle, Etensor carpi ulnaris muscle, Flexor carpi ulnaris muscle and Latissimus dorsi muscle.

Next, analysis was performed by extracting muscle potential during standing and sitting movement from measuring result with reference to middle trochanter and trunk angle and calculating value of interal during movement. Table 1 shows the ratio of value integral with welfare device to one without device. Also, we compared moving distance of median point between using welfare device and not. Table 2 shows the comparison results in a manner similar to Table 1.

B. Evaluation by using NIRS

As common result of all subjects, oxy-Hb tended to increase during task and to decrease in resting state. Therefore, it was thought that change of hemoglobin density due to task was measured. Fig. 6 shows trend of the channel in which significant different was shown. Analysis was performed via one-sample t-test [5,6,7,8,9] by a method similar to previous researches [5,6,7,8,9]. In this analysis, it was necessary to remove other than change of blood flow due to fatigue. So, our method was mainly focused on resting state to compare with the 1st trial and another trials of brain activity.

In task1 1 and 2, each of sample data for analysis was 4 seconds after the task(Fig. 4). In task 3, sample data was 4 seconds during task.(Fig. 5)

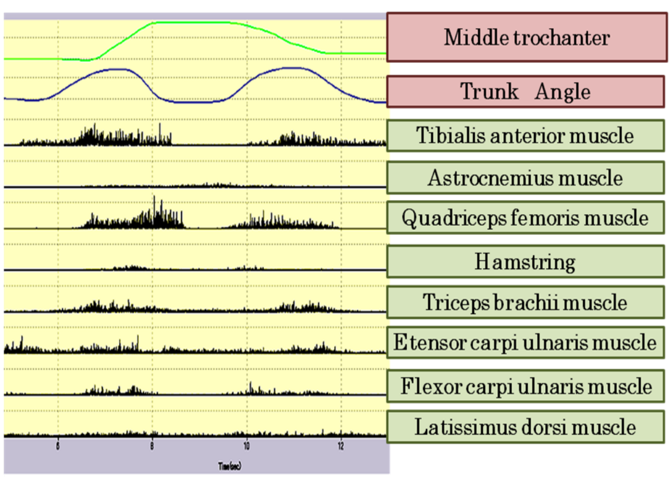


Figure 3. Result of 3D Motion Analysis and EMG

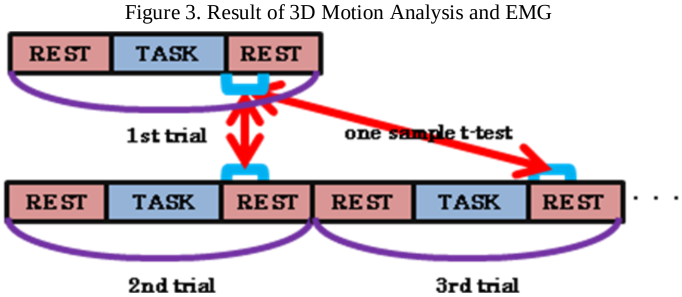


Figure 4. T-test of sample data of task1 and 2

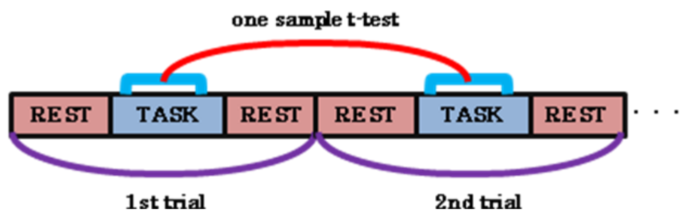


Figure 5. T-test of sample data of task3

TABLE I. Comparison of Integral EMG

muscle	region	Subject1	Subject2	Subject3
Standing	Tibialis anterior muscle	0.37	0.49	0.64
	Astrocnemius muscle	0.83	0.78	0.97
	Quadriceps femoris muscle	0.66	0.36	0.81
	Hamstring	1.90	0.50	1.07
	Triceps brachii muscle	1.07	3.34	1.01
	Etensor carpi ulnaris muscle	1.08	1.31	0.96
	Flexor carpi ulnaris muscle	1.07	0.89	0.85
	Lattissimus dorsi muscle	0.98	0.87	1.20
Sitting	Tibialis anterior muscle	0.50	0.59	0.80
	Astrocnemius muscle	1.01	0.92	0.94
	Quadriceps femoris muscle	0.49	0.57	0.85
	Hamstring	2.16	1.60	0.96
	Triceps brachii muscle	0.89	0.96	1.07
	Etensor carpi ulnaris muscle	0.79	0.89	0.86
	Flexor carpi ulnaris muscle	0.79	0.86	0.95
	Lattissimus dorsi muscle	1.16	1.18	0.93

TABLE II. Comparison of change in medial point

	Subject1	Subject2	Subject3
Sitting	0.89	1.03	0.90
Standing	1.00	0.84	1.08

In the t-test of the same task, we performed t-test with first time trial and other trial which was from second times to thirty times, and examined relationship the number of trials and significant differences.

In task 1, significant different could be found from the about 10th trials. Fig. 9 show region confirmed significant difference.

In task 2, significant different could be found from the about 10th trials too. Fig. 10 show region confirmed significant difference.

Next, we performed t-test with case of standing position (task 1) and half-crouch position(task 2). In this analysis, significant different could be found at prefrontal area(14ch, 17ch, 28ch and 32ch). Fig. 11 show region confirmed significant different.

Also, two types of motion which was sitting and keeping a half-crouching position were repeated alternatively in task 3. At first, we performed t-test using 4 seconds during first trial and 4 seconds during other trials, which were from second to fifteenth in same position. Regarding to analysis result using sample data during sitting position and half-crouching position, there were significant different at Prefrontal area. Fig. 12 confirms significant difference.

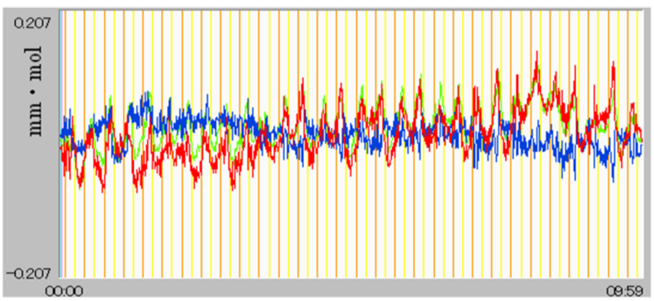


Figure 6. Measuring Result of Task1

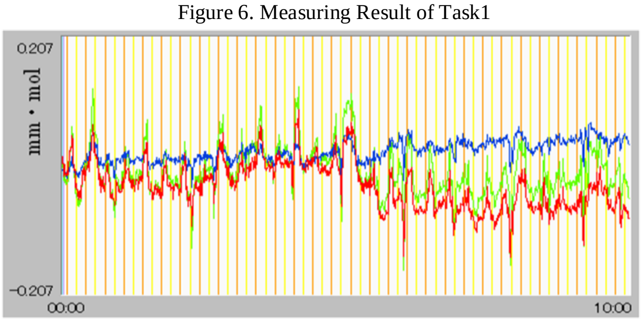


Figure 7. Measuring Result of Task2

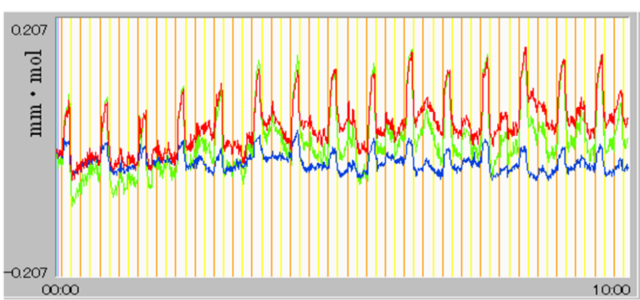


Figure 8. Measuring Result of Task3

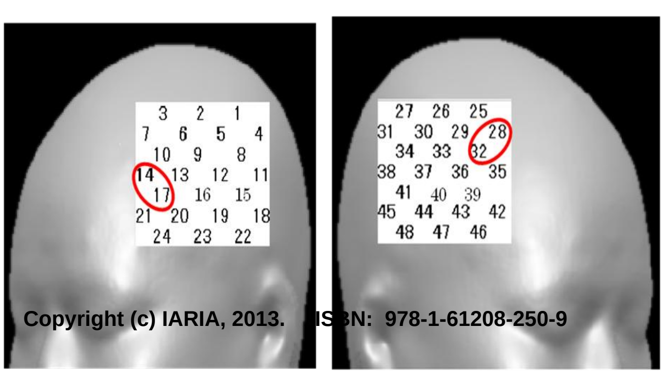


Figure 9. Signififant Difference of task1

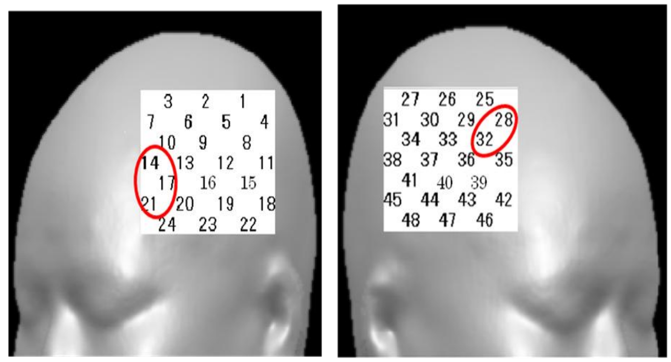


Figure 10. Significant difference of task2

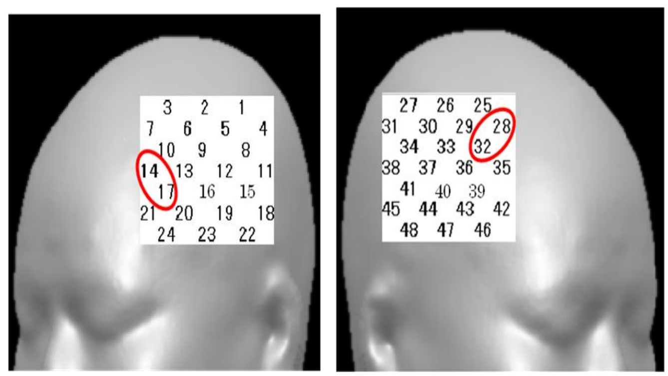


Figure 11. Significant Difference of task 1 and 2

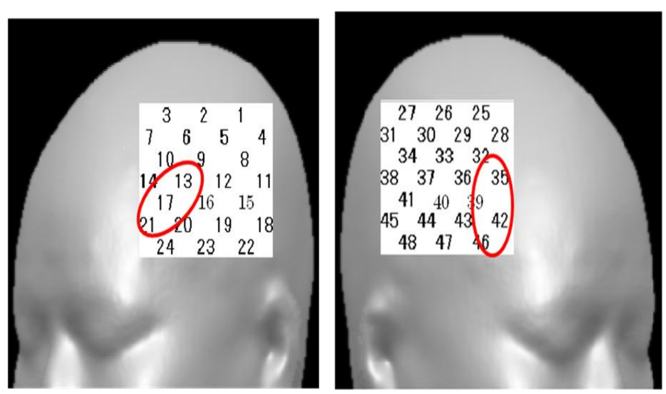


Figure 12. Significant Difference in sitting position

IV. Discussion

1) Evaluation by using 3D Motion Analysis and EMG

From analysis result, it was shown that value of integral was decreased by using assistive apparatus for transfer. Especially, there was remarkable decrease in value of integral at tibialis anterior muscle, quadriceps femoris muscle. On the other hand, it was shown to be minor decrease in one at upper limb and muscles of the back. Also, moving distance of barycentric position was decreased by the use of welfare device.

On the ground of this result, it was thought to be due to difference in height between chair and seating face of welfare device. Therefore, it was thought that the use of assistive apparatus is useful to lighten burden on lower limb. Thus, it's contemplated that muscle load during standing and sitting movement was decreased and reduced centroid fluctuation to lower the possibility of turnover.

Even if subjects performed daily movements of standing and sitting with the use of assistive equipment, it was shown that the integral of muscle potential and distance of centroid change was decreased. Therefore, it was proved that there is the possibility of evaluation of daily performance except for movement with welfare device.

2) Evaluation by using NIRS

In this experiment, we tried to measure quantitatively the physical and psychological strain on the basis of brain activity. Also, we think that brain activity disclose human cognitive including non-verbal. As a result, it was shown that there were differences at brain activity due to number of trials and postural. In this time, analysis was performed via one-sample t-test using sample of brain activity in resting state during task or after task. Hence, analysis method was to remove disturbance such as body motion and angular variation of neck to the extent possible although there was the possibility to measure skin blood flow. Therefore, it was thought that strain due to tasks was quantitatively measured by being recognized significant differences

Also, in previous research, it was reported to decrease in activity in the brain around #10, 11 [10], as the result of measuring brain activity during Advanced Trial Making Test using PET [11]. Therefore, this result came out in support of previous research in no small part.

Of course, it is necessary to increase number of subject at the present stage. In addition, there are problems associated with experiment, number of subject, method and measured region. However, in terms of being recognized significant differences at brain activity due to movement, it was thought to show useful result in evaluating quantitatively daily movements.

V. CONCLUSION AND FUTURE WORK

In this experiment, our purpose was to evaluate quantitatively physical load with focusing on standing and sitting movement which was part of daily movements using 3D motion analysis system and EMG.

As the result, it was shown that the integral of lower-limb muscle, such as tibialis anterior muscle and gastrocremius muscle, significantly decreased by the use of welfare device.

Also, it was reported that there is a positive correlation between anteversion angle of body trunk and movement duration in previous research [12]. But, our experiment method was to estimate the possibility of falling in rising from a sitting position by calculating moving distance of median point. And, it was confirmed that the possibility of falling was decreased by using device.

Next, we tried to measure physical and psychological load quantitatively on the basis of brain activity. And there were significant differences due to number of trials, holding position. In this experiment, analysis method was to remove disturbance such as body motion and angular variation of neck to the extent possible by using measurement result in resting state as sample. Therefore, it was thought to show useful result in evaluating quantitatively load due to movement task by being recognized difference in brain activity caused by number of trials, substance of task and holding position.

Main purpose in this study is to evaluate physical load and fatigue quantitatively. So, we tried to evaluate change of muscle load due to difference of motion by simultaneous measuring with 3D motion analysis System and EMG quantitatively.

However, evaluation of psychological load is necessary, too. In terms of using welfare device, prolonged use must be taken into account. In this case, it is important to consider not only physical load but also psychological load due to prolonged use from standpoint of developing welfare device and keeping up surviving bodily function.

Also, in previous research, separation between physical and psychological load has been performed. But, our view is that there is correlation with physical and psychological load. So, we tried to measure psychological load including physical one based on brain activity and quantitatively evaluate both load.

For the future, our aim is to establish method of discussing useful of welfare device by evaluating load involved in other daily movements with increasing number of subjects.

ACKNOWLEDGMENT

This research was supported as follow.There dimensional motion analysis and commodious room were provided from nac Image Techonogy Inc. EMG and analysis software were provided from KISSEI COMTEC. NIRS was provided from SHIMADZU COPPORATION. Developing welfare device was provided from IDEA SYSTEM CO., LTD. We are deeply grateful to them.

This study contributes to become the basis for one of theme of s-innovation program in Japan Science and Technology Agency which was named “Development Fatigue-reduction Technology for Social Contribution of Aged Person and Establishment System for Evaluation.

REFERENCES

[1] Y. Shinoda, “Considertion of feature extraction based on center of gravity for Nihon Buyo dancer using motion capture system,” SICE Annual Confference 2011, Tokyo Japan, pp. 1874-1878.

[2] Y. Yamaguchi, A. Ishikawa, and Y.Ito, “Development of Biosignal Integration Analysis System for Human Brain Function and Behavior,” Organization for Human Brain Mapping 2012, China.

[3] H.Inoue, S. Shimizu, N. Takahashi, H.Nara, and T. Tsuruga “Fundamental Study for Evaluation of the Effect due to Exercise Load” , Assistive Technology, Bio Medical Engineering and Life Support 2011, Japan.

[4] E. Watanabe, Y. Yamashita, Y. Ito, and H. Koizumi, ”Non-invasive functional mapping with multi-channel near infra-red spectroscopic topography in humans,” Heurosci Lett 1996, Feb 16, 205(1), 41-4.

[5] N. Takahashi, S. Shimizu, Y. Hirata, H. Nara, F. Miwakeichi, N. Hirai, S. Kikuchi, E. Watanabe, and S. Kato, “Fundamental Study for a New Assistive System during Car Driving,” Proc. of

 ISBN: 978-1-61208-250-9

International Conferrence on Robotics and Biomimetics, DVD-ROM, 2010, Tenjin, China.

[6] N. Takahashi, S. Shimizu, Y. Hirata, H. Nara, H. Inoue, N. Hirai, S. Kikuchi, E. Watanabe, and S. Kato, "Basic study of Analysis of Human Brain Activities during Car Driving," the 14th International Conferrence on Human-Computer Interaction, 2011, USA.

[7] S. Shimizu, N. Takahashi, H. Nara, H. Inoue, and Y. Hirata, "Fundamental Study for Human Brain Activity Based on the Spatial Cognitive Task," the 2011 Internatinal Conference on Brain Informatics-BI 2011, China.

[8] S. Shimizu, N. Takahashi, H. Nara, H. Inoue, and Y. Hirata, "Basic Study for Human Brain Activity Based on the Spatial Cognitive Task," The Third International Conference on Advanced Cognitive Techonologies and Applications, 2011, Italy.

[9] S. Shimizu, N. Takahashi, H. Inoue, H. Nara, F. Miwakeichi, N. Hirai, S. Kikuchi, E. Watanabe, and S. Kato, "Basic Study for a New Assitive System Based on Brain Activity associated with Spatial Perception Task during Car driving," Proc. International Conferrence on Robotics and Biomimetics, 2011, Thailand.

[10] Y. Watanabe,"Molecular/neural mechanisms of fatigue, and the way to overcome fatigue," Folia Pharmacological Japonica, vol. 129, pp. 94-98, 2007.

[11] H. Kuratsune, K. Yamaguti, G. Lindh, B. Evengard, G. Hagberg, K.Matsumura, M. Iwase, H. Onoe, M. Takahashi, T. Machii, Y.Kanakura, T. Kitani, B. Langstrom, and Y. Watanage,"Brain Regions Involved in Fatigue Sensation: Reduced Acetylcarnitine Uptake in to the Brain," Neuroimage, vol. 17, pp. 1256-1265, November 2001.

[12] K. Maruta,"The influence of Seat Angle on Forward Trunk Inclination During Sit-to-Stand," Jounal of Japanese Physical Therapy Association, vol. 31, No.1, pp. 21-28, 2004.

Haptic Manipulation of Objects on Multitouch Screens:

Effects of screen elevation, inclination and task requirements on posture and fatigue

Samantha Scotland, Shwetarupalika Das, Thomas Armstrong, Bernard J. Martin
Industrial & Operations Engineering
University of Michigan
Ann Arbor, U.S.A
e-mail: sscotlan@umich.edu, shdas@umich.edu, tja@umich.edu, martinbj@umich.edu

***Abstract*—Application of multi-touch screen interfaces is desired by manufacturers of measuring instruments to follow the current trend of consumer products. This study is a first approach to determine the influence of screen location and orientation on upper limb movements and posture. Neck and wrist posture, as well as finger movements and subjective perception of discomfort were evaluated in simple tasks simulating the placement and scaling of objects using a multi-touch interface placed on a computer screen of the same size. The results show that wrist and neck postures are generally affected by an interaction between screen height and inclination. They also suggest that precision requirements associated with measuring instruments may not be compatible with simple transfer of manipulation methods from consumer products.**

Natural gestures; finger movements; wrist posture; discomfort

I. INTRODUCTION

Natural user Interface technologies based on multi-touch displays have been successfully implemented in a variety of consumer devices. This success has lead to the adoption of these technologies for work applications. However, most work instruments are designed to sit on a bench or cart, have vertically mounted screens and may require precise manipulations. Gestures such as the “pinch”, used to zoom in and out, may require awkward wrist and upper limb postures when performed on a vertically mounted screen. Posture and repetition are known risk factors [1-3] commonly associated with localized muscle fatigue [4-6] and upper limb musculoskeletal disorders [7-8]. Furthermore, movement precision required to set specific values of gesture-controlled functions may increase repetition and muscle co-contraction activities. Hence, simply using the methods designed for mobile devices on vertically mounted screens may create rather than solve problems.

The present study investigated the influence of screen height (elevation) and inclination (relative to the vertical direction) on wrist posture and finger movements during object manipulation tasks using simple natural gestures. The level of required precision varied between tasks.

II. MATERIAL AND METHODS

A. *Participants*

Ten right-handed young adults aged 20-31 years (5 males, 5 females) participated in the experiment as volunteers. All were free from upper limb neurological and musculoskeletal disorders that might have impaired their performance. Their visual acuity was 20/20 or better, with or without correction.

B. *Experimental setup and conditions*

An instrument screen was simulated on a computer screen, onto which a matched size multi-touch screen was adapted. LED markers were placed on the thumb and index fingers, the wrist, elbow and shoulder of the right limb and head to record their movements with an active motion capture system (Optotrack™). The participants were facing the screen seated on a comfortable office chair adjusted to their anthropometry to obtain a 90° knee angle with feet flat on the ground. Three screen heights (1.25, 1.5 and 1.75 seated height- measured to the bottom of the screen) combined with three screen inclinations (10, 15, 20° relative to the vertical direction) were tested in random order. The tasks to be performed simulated real instrument operations. They included adjusting both the vertical and horizontal position of an object, scaling the horizontal and vertical dimensions of the object to specified values and switching the primary object to be

 ISBN: 978-1-61208-250-9

manipulated. Numerical values and grid scales were displayed on the screen.

C. *Procedure*

The duration of each of the five tasks varied between 30 s and 2 min. A brief 10-20 s rest separated two consecutive tasks. A longer 2 min rest was provided between conditions while the experimenter adjusted the screen. Between trials, participants were allowed to relax their arm on the arm rest and between conditions they could stand or move away from the work station. The motion signals associated with each marker were sampled at 100Hz.

D. *Measures*

Discomfort was evaluated on a category scale (0= no discomfort, 1= barely perceptible, …, 5 = extreme pain) for each of the following body parts: finger, wrist, elbow, shoulder, neck, back). The magnitude of selected angular displacements (wrist flexion-extension; wrist pronation-supination; neck flexion-extension) and thumb-index finger "pinch" opening indicated by the maximum distance between the respective finger tips distance were computed from the recorded motion of the respective markers. Observations by the experimenter about task execution and comments from participants were recorded for each experiment for association with quantitative data

III. RESULTS

A. *Discomfort rating*

No major discomfort or fatigue was reported in the context of our experiment. Ratings were rather low; however neck and shoulder discomforts were higher for the 20° screen inclination and were also generally higher for the highest screen elevation. The average peak discomfort corresponded to a value of 2 on the 0-5 scale and discomfort tended to increase with the chronologic order of experimental conditions.

B. *Wrist, head and finger movements*

Wrist extension increased primarily with screen verticality (low inclination), as illustrated in Fig.1. The order of screen height indicates that wrist extension tend to be affected also by screen elevation. However, a higher elevation may reduce wrist extension induced by screen inclination, as indicated by a significant elevation-inclination interaction.

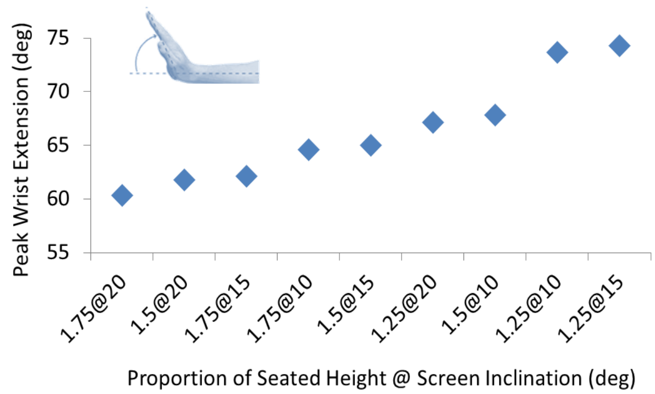


Figure 1. Wrist extension as a function of experimental conditions (screen elevation and inclination).

Wrist rotation tended to increase primarily as a function of screen height, as illustrated in Fig.2.

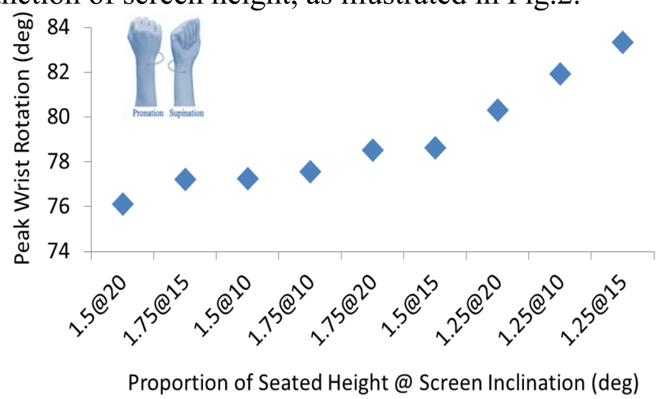


Figure 2. Wrist rotation as a function of experimental conditions.

Neck flexion seems to be higher for the lower screen inclinations (more vertical), particularly for the lower screen elevation as illustrated in Fig.3.

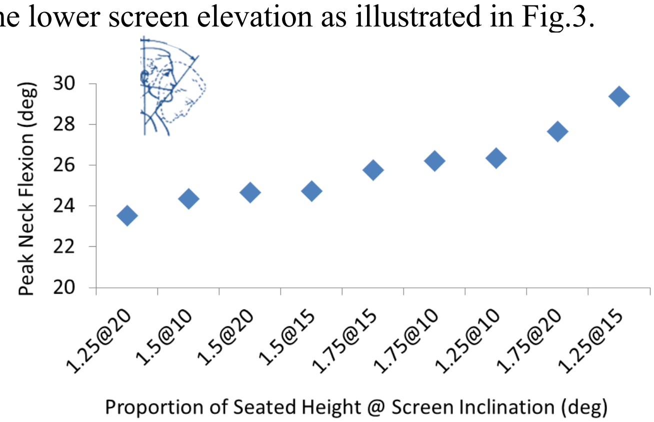


Figure 3. Neck flexion as a function of experimental conditions.

Index finger displacement relative to the thumb ranged from 11.75 cm to 12.9 cm; however,

 ISBN: 978-1-61208-250-9

movement patterns did not appear to be associated with specific experimental conditions (Fig.4). Furthermore, task duration and thus movement repetition increased significantly (up to 5 fold) when position and scaling precision were required.

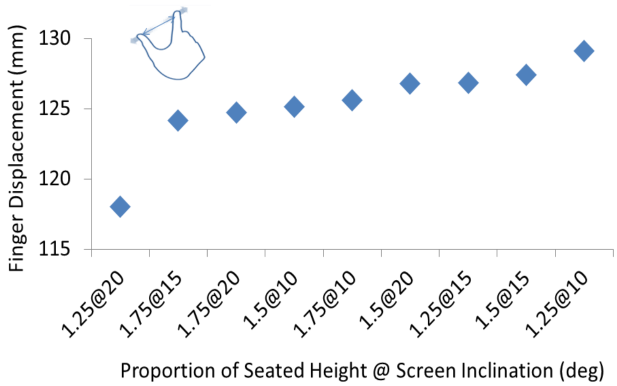


Figure 4. Thumb-index finger maximum distance as a function of experimental conditions.

C. *Survey*

Comments from the participants indicated that object placement and size scaling were very/extremely difficult to achieve when high precision was required (100% of participants). In addition, a delay between finger movements and object movements, as well as parallax associated with the distance (although only a few mm) between the touch screen and computer screen were frequently cited (70% of participants) as factors affecting performance.

IV. DISCUSSION

The results indicate that discomfort was not an issue in the context of our experiment; however, the influence of screen elevation and inclination on wrist posture was significant and showed expected interactions. In addition, delay and the need for precise manipulation contribute to an increase in the number of movements necessary to complete the tasks. Furthermore, multi-touch operations, such as size adjustment, frequently required hand repositioning to re-initiate a gesture.

Discomfort (or muscle fatigue) was moderate in the present context; however the tendency of discomfort to increase with time, as indicated by a sequential order effect suggests that prolonged "manipulations" may lead to the development of fatigue, even if finger forces required to move the fingers are very low [9]. In this experiment, the participants could rest their arm on the armrest at anytime. They could also use the bottom edge of the screen as a wrist support (1) in between tasks, (2) when manipulation of the object did not necessitate vertical displacement of the wrist, or (3) when inclination of the screen permitted such posture / use of support reaction forces. No specific instruction was provided; hence the participants adopted strategies associated with muscular effort minimization, as is the case in posture control [10]. However, these "support/anchoring" strategies are highly dependent on the availability of support and environmental context in which the instrument is used. Furthermore, the participants were not time constrained and benefited from rest periods while the screen position was adjusted for each condition.

Not surprisingly, wrist extension and rotation were influenced by both screen elevation and inclination. In addition, it may be assumed that influences from task specificities are very likely since different gestures (finger movements) require wrist and arm coordination as a function of the direction of the movement. A more detailed analysis based on a larger number of repetitions for each task would be necessary to determine more precisely optimal screen location to minimize large deviations from neutral joint postures. Nevertheless, the present results indicate that touch screen manipulations on vertically mounted instruments may induce significant wrist deviations, which are considered a risk factor, especially when sustained or combined with high repetition [7]. Furthermore, repetition seems to be promoted by accuracy requirements, at least when simply duplicating manipulation methods used to operate hand held devices.

Variations in neck flexion with screen elevation and inclination have been associated with maintaining the line of sight in computer tasks [11-13]. However, in the present context, these variations may reflect an optimization of head orientation to improve visual guidance of hand movements [14].

Overall, postural adjustments seem to be driven by the environmental conditions and the manual task to be performed, which differ from classic visual capture of information from a video screen.

ACKNOWLEDGMENT

The authors would like to thank E. Claxton and C. Woolley for their technical assistance.

REFERENCES

[1] C. Fransson-Hall, S. Byström, and A. Kilbom, “Characteristics of forearm-hand exposure in relation to symptoms among automobile assembly line workers,” Am J Ind Med, vol. 29, Jan. 1996, pp. 15–22, doi: 10.1002/(SICI)1097-0274(199601)29:1<15::AID-AJIM3>3.0.CO;2-Y.

[2] USA, US National Institute for Occupational Safety and Health, Department of Health and Human Services (DHHS), “Musculoskeletal disorders and workplace factors: a critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back,” B. P. Bernard, Ed. Cincinnati, OH: NIOSH, 1997, Pub No 97-141.

[3] B. Silverstein, E. Welp, N. Nelson, and J. Kalat, “Claims incidence of work-related disorders of the upper extremities: Washington state, 1987 through 1995,”Am J Public Health, vol. 88, Dec. 1998, pp. 1827-1833.

[4] S. Byström and C. Fransson-Hall, “Acceptability of intermittent handgrip contractions based on physiological response,” Hum Factors, vol. 36, no. 1, Mar. 1994, pp. 158–171, doi: 10.1177/001872089403600110.

[5] M. Gerard M, T. Armstrong, J. Foulke, and B. J. Martin, “Effects of key stiffness on force and the development of fatigue while typing,” Am Ind Hyg Assoc J, vol. 57, Sept. 1996, pp. 849–854, doi: 10.1080/15428119691014549.

[6] D. E. Adamo, M. Khodaee, S. Barringer, P. W. Johnson, and B. J. Martin, “Low mean level sustained and intermittent grip exertions: influence of age and fatigue on recovery,” Ergonomics, vol. 52, no. 10, Oct. 2009, pp. 1287-1297, doi:10.1080/00140130902984935.

[7] T. J. Armstrong, P. Buckle, L. J. Fine, M. Hagberg, B. Jonsson, et al. “A conceptual model for work-related neck and upper-limb musculoskeletal disorders,” Scand J Work Environ Health, vol. 19, Apr. 1993, pp. 73–84.

[8] D. Chaffin, G. Naderson, and B. J. Martin, Occupational Biomechanics, 4th ed., New York, NY: John Wiley & Sons, Inc., 2006.

[9] G. Shin and X. Zhu, “User discomfort, work posture and muscle activity while using a touchscreen in a desktop PC setting,” Ergonomics, vol. 54, no. 8, Aug. 2011, pp. 733-744, doi: 10.1080/00140139.2011.592604.

[10] A. D. Kuo, “An optimal control model for analyzing human postural balance,” IEEE T Bio-Med Eng, vol. 42, no.1, Jan. 1995, pp.87-101, doi: 10.1109/10.362914.

[11] D. Colombini, E. Occhipinti, C. Frigo, A. Pedotti, and A. Grieco, “Biomechanical, electromyographical and radiological study of seated postures,” in The Ergonomics of Working Postures: Models, Methods and Cases, N. Corlett, J. Wilson, and I. Manenica, Eds. London: Taylor & Francis, 1986, pp. 331- 344.

[12] N. Hamilton, “Source document position as it affects head position and neck muscle tension,” Ergonomics, vol. 39, no. 4, Apr. 1996, pp. 593-610, doi: 10.1080/00140139608964483.

[13] M. B. G. Villanueva, M. Sotoyama, H. Jonai, Y. Takeuchi, and S. Saito, “Adjustments of posture and viewing parameters of the eye to changes in the screen height of the visual display terminal,” Ergonomics, vol. 39, no. 7, Jul. 1996, pp. 933-945, doi: 10.1080/00140139608964515.

[14] K. H. Kim, B. Gillespie, and B. J. Martin, “Head movement control in visually guided tasks: postural goal and optimality,” Comput Biol Med, vol. 37, no. 7, Jul. 2007, pp. 1009-1019, doi: 10.1016/j.compbiomed.2006.08.019.

Sliding Raised-Dots Perceptual Characteristics

Speed Perception or Dot Count

Yoshihiko Nomura
Graduate School of Engineering
Mie University
Tsu, Japan
nomura @ mach.mie-u.ac.jp

Syed Muammar Najib Syed Yusoh
Graduate School of engineering
Mie University
Tsu, Japan

Kazuki Iwabu
Graduate School of engineering
Mie University
Tsu, Japan

Ryota Sakamoto
Mie University Hospital
Tsu, Japan

***Abstract*— The authors have studied new mode of cutaneous sensation characteristics on finger pads toward an accurate physical-line presenting computer-human interface. A series of psychophysical experiments were carried out on motion perception characteristics with the raised dots sliding on finger pads. As a result, it was found that there were two kinds of modes with the perception: (1) dot-counting mode, (2) raised-dots sliding-speed perceptual mode. The first mode worked in the long period of dot spacing and in the low sliding speed conditions, and showed such a high accuracy as a proprioceptive-sensation-based fingertip motion perception. The second mode worked in the short period or in the high speed conditions, and suffered ill performance similar to sliding flat surface.**

Keywords; cutaneous sensation; fingerpad; sliding; raised dots; counting; speed

I. Introduction

We can accept such geometrical information of objects as the lengths and directions of line-segments via our haptic sensations. Therefore, the haptic sensations have been expected as an alternative for visually impaired persons to create geometrical mental images. In the haptic sensations, there are two important factors: the sensation factor and initiative factor. As for the sensation factor, not only the cutaneous sensation but also the proprioceptive one was employed in perceptual processes. On the other hand, as for the initiative factor in the hand motion activation, there are such two schemes as the passive and active schemes.

The cutaneous sensation is a principal cue in hand-stationary passive touches, and is also used as a supplementary cue to proprioceptive one in hand-moving passive/active touches. Therefore, we can say an importance of investigating human cutaneous sensations for understanding the haptic perceptual mechanism.

Thus, there have been studied many cutaneous-sensation-based perceptual characteristics. For example, slip length perceptual characteristics of moving flat surface were studied in [1]~[4]. Under the active touch condition, Hollins et al. reported a perceived-length formula as a function of velocity and duration [1]. Armstrong et al., focusing on the duration time, suggested that the radial–tangential anisotropy in the length perception can be explained by temporal differences in exploratory movements [2]. In these years, the passive length perception scheme has been studied in detail. For example, Terada et al. found a complimentary characteristic between the cutaneous and proprioceptive sensations and proposed some formula representing the perceived length as the function of velocity dependence and actual-length [3]. Wouter et al. reported that the cutaneous condition (Cu) was much inferior to a proprioceptive condition (Pr) and a cutaneous-combined proprioceptive condition (CuPr). In the CuPr-condition, the cutaneous sensation contributes a little to the Pr-condition [4]. Introducing another important aspect of movement direction, Yusoh et al. have also presented other formulas representing the perceived length as the function of velocity dependence and actual-length [5].

As for the dots-counting sensation that was an important factor in this work, there were few researches except for those on vibrotactile sensations. For examples, papers of [6~9] took up the perception of single site vibration. However, we can't find such similar studies on moving-raised-dots produced stimuli as in this work, and papers of [10~12] took up the perception of pattern-like vibration..

Raised dots have been used for Braille, i.e., dotted letters and for tactile-graphics since they have been considered to be distinctive and informative stimuli for representing geometrical information. Therefore, the raised dots were expected to contribute for new physical-line presenting interfaces: for example, forming raised dots on actively rotating wheel surface, by embedding the wheel, we could make up novel active mouse interface.

Thus, the raised dot was expected to be a promising tactile presentation framework. In this study, paying notice to the sliding-raised dot pathway perception, the authors have studied cutaneous sensation characteristics on finger pads in the passive touch scheme.

In this paper, Section 2 introduced such methodology of experiments as the apparatus and procedures. Next, comparing some related existing works, Section 3 presented experimental results with the sliding length perceptual characteristics for the short/long period dot specimens. Here, it was referred that a transition of the sliding dot perceptual mode from dot counting to sliding speed perception occurred even in the long period dot specimen. Finally, Section 4 closed this paper by summarizing important findings.

II. EXPERIMENTAL METHOD

A. Apparatus

Figure 1 shows the experimental apparatus. The length, width, and height of table base are 510×500×610mm, respectively. It consisted of a power supply, a controller, an acrylic plate with a hole, a specimen plate, a linear actuator (IAI-ICSA series) and a rotation board. The acrylic plate was mounted on the top of the flat plate. A hole was opened at the center of the acrylic plate. The specimen plate, to the top surface of which paper specimen adhered, was attached on the top surface of the linear actuator and it could be moved in both the directions within the 300mm range. The gap between the acrylic and specimen plates was set at 6mm. The linear actuator was, furthermore, mounted to the top of the rotation board so that the linear actuator could be rotated and be fixed at a direction within 360° with an interval of 22.5°.

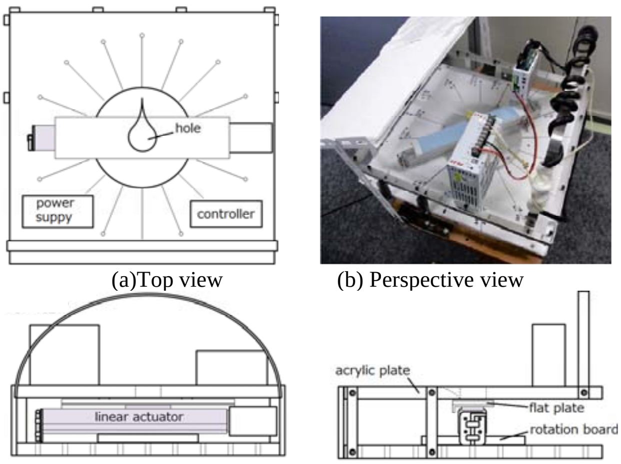


Figure 1. Experimental apparatus: a specimen was moved by a linear actuator on a rotation board.

We introduced two kinds of specimens, on the surfaces of which a raised dot pathway was formed so as to be aligned in straight: one specimen had a 20.1mm-period of raised dots and the other did a 2.5mm-period. The raised dots were 1.5mm in diameter and 0.4mm in height (see Figure 2). In the case of 20.1mm period dot, the dot appeared one-by-one on the contacting fingerpad. On the other hand, in the other case of 2.5 mm period, multiple dots contacted with the fingerpad. This gave birth to an important difference in perceptual mode between the two periods. In the case of 20.1mm period, we could perceive each of the sliding raised dots one-by-one with the finger pad cutaneous sensation and could count the number. Meanwhile, in the other case of 2.5mm period, we no longer could not count them but, as an alternative, perceived sliding speed of the raised dot. The characteristics for both of the perceptual modes were discussed in the experiments.

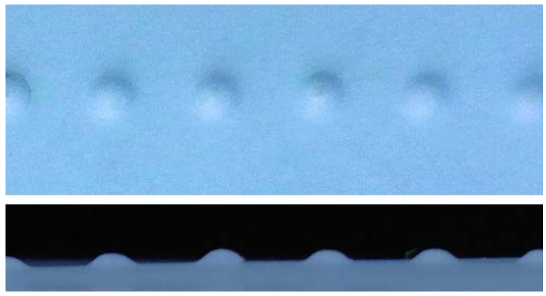


Figure 2. Specimen (upper: top view. lower: front view).

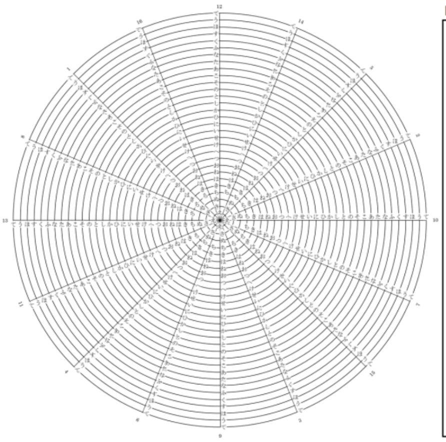
B. Procedures

1) Conditions

The line lengths being presented were 50, 75, 100, 125, 150mm, and the directions were set at 0 to 360° with an interval of 22.5°. Therefore, the number of presented lengths was five and that of the directions was 16. In the case of 20.1mm period, the lengths, 50, 75, 100, 125, 150mm, corresponded to 2, 3, 4, 6, and 7 dots, respectively. The 80 line segment patterns by 5 lengths and 16 directions were presented in pseudo random orders. All the experiments were carried out with the speeds of 25, 50, 75, 100, 125, 150, 175 and 200mm/s. Five subjects aged 22 to 57 years old participated in the experiments.

2) Task

In the case of 2.5mm period dots, the subjects were informed to perceive both the sliding length and direction during the dotted paper specimen sliding on his fingerpad. They were also asked to make oral reports of the perceived length and direction with the character-represented codes: the codes were written on an answer board being showing concentric circles, of which radii were 5 to 200mm with a 5mm interval (see Figure 3). In the other case of the 20.1mm-period-dot, the subjects were informed to count the number of the dots, and also to perceive the sliding direction as one of the 16 candidate directions during the dotted specimen sliding on their finger pad. Just after the slide having been finished, the subjects made oral reports of the frequency of counted dots, and the perceived direction.



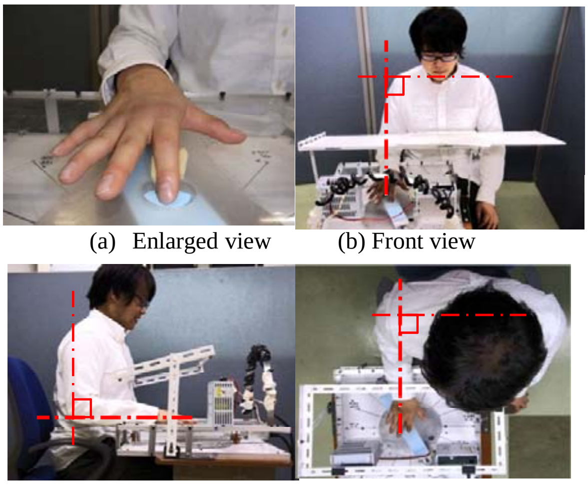
Direction		Length			
10	0°,360°	え	5	か	80
2	22.5°	さ	10	し	85
5	45°	ぬ	15	と	90
14	67.5°	ち	20	の	95
12	90°	き	25	そ	100
16	112.5°	は	30	こ	105
1	135°	ね	35	あ	110
8	157.5°	お	40	た	115
13	180°	つ	45	な	120
11	202.5°	へ	50	ふ	125
4	225°	け	55	く	130
6	247.5°	せ	60	す	135
9	270°	い	65	ほ	140
3	292.5°	に	70	う	145
15	315°	ひ	75	て	150
7	337.5°				

Figure 3. Answer board: subjects chose a pair of words to answer the perceived direction and length from a word set shown in the answer board

3) Procedures

Figure 4 shows some experimental views. Subjects were seated in a chair, setting their right forearm on the acrylic plate. During experiments, subjects were instructed to relax, and to focus on to perceive the presented linear slide motions. The subjects put their index fingerpad on the surface of a flat specimen (Figure 4(a)) where the subject arm angle was set at 90° (Figure 4(b)-(d)), and the arm direction was set to horizontal to table base (Figure 3(c)). The subject aligning his shoulder in the frontal plane and holding his upper body not twisted at the waist was placed at the position so that his right arm was parallel to the sagittal plane direction and so that his index fingertip was located at the hole in the acrylic plate (Figure 4(d)). A white noise sound was applied to the subjects via headphones for masking any sound cues on the spatial perception. In an arbitrary waiting time, after the subject put his fingerpad on a specimen, the specimen started to move, and stopped at a given point with approximate rectangular velocity pattern: by using the linear actuator, the specimen was accelerated with 3 m/s^2, then, it was moved with constant speed after reaching at the predetermined speed, and it was finally decelerated by 3 m/s^2.

The direction of the linear actuator movement was set by manual operation each time. The velocity and direction were pseudo-randomly presented for each of the subjects. It took about 30 minutes for the whole experiment per a subject.



(a) Enlarged view (b) Front view

(c)Side view (d) Top view

Figure 4. Experimental view showing the posture of a subject

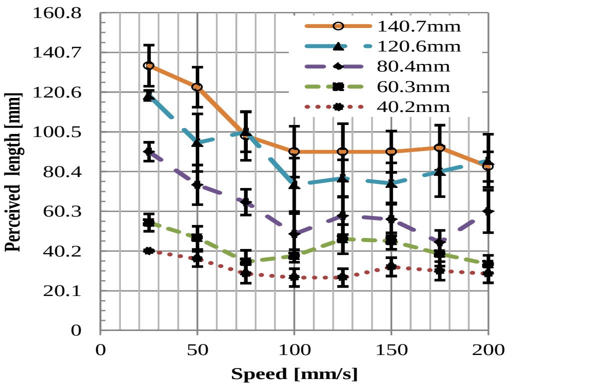
III. Experimental Results

A. From Dot Counting to Sliding Speed Perception

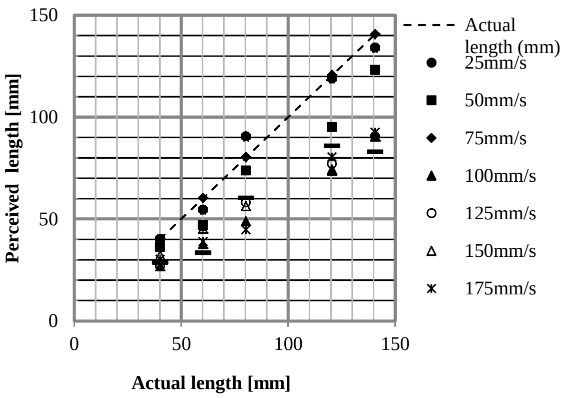
Figure 5(a) shows an experimental result for the 20.1 mm period raised dot specimen. In this figure, counted dot numbers were inverted into length dimension by multiplying 20.1. In this case, a speed-caused fewer-counting effect in means (it is rephrased as a speed-caused foreshortening effect from the viewpoint of the length dimension) can be seen clearly. However, it should be noted that there can be seen little mean errors in the case of the 25 mm/s velocity. This characteristic will be useful to develop the motion presenting interfaces in the future. To make clear actual and counted dot number relationship, changing the parameters from the lengths to speeds, we got Figure 5(b) from Figure 5(a). From this figure, we can see the length-related foreshortening effect, especially, in high speed conditions. As a whole, it can be said that, the more the velocity/length is, the bigger the mean error grows. On the other hand, as for random errors, the longer the actual length was, the more the STD of the actual dots number increased.

 ISBN: 978-1-61208-250-9

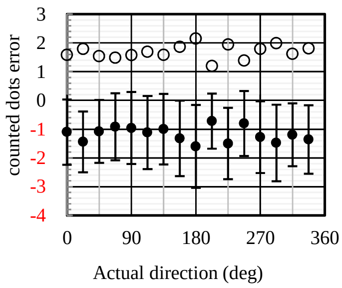
Figure 5(c) shows a directional variation in dot counting. We couldn't clearly find any directional characteristic.



(a) Means of perceived lengths (symbols), together with their standard errors(error bars): actual lengths were taken as a parameter.



(b) Means of perceived lengths: speed was taken as a parameter.



(c) Directional characteristic of perceived dot number: filled circles, error bars, and open circles were mean errors, standard devitions, and RMSE

Figure 5. Experimental results with the 20.1 mm period raised dots

The other experimental result is shown in Figure 6 for the 2.5 mm period raised dots.

For each of the 20.1 mm and 2.5 mm period dot experiments, we formulated a model for the perceived length l_{est_perc} by using a power function of actual length l_{act} and velocity v.

$$l_{est\_perc} = \alpha\, v^{\beta}\, l_{act}^{\gamma}. \qquad (1)$$

The power function model is a variation of Stevens' law, and it represents contraction effects along with increased speed and length. In the ideal case that there are no such effects, the exponents with respect to l_{act} and v, i.e., β and γ, are respectively 1 and 0, and the proportional coefficient is 1. When estimating the unknown parameters, we took logarithms of the both sides of Eq. (1). That is,

$$\ln l_{est\_perc} = \ln\alpha + \beta \ln v + \gamma \ln l_{act} . (2)$$

For the 20.1 mm period experiment, after converting the counted dot numbers into lengths, we got

$$l_{est\_perc} = 2.2\, v^{-0.20}\, l_{act}^{0.95}. \qquad (3)$$

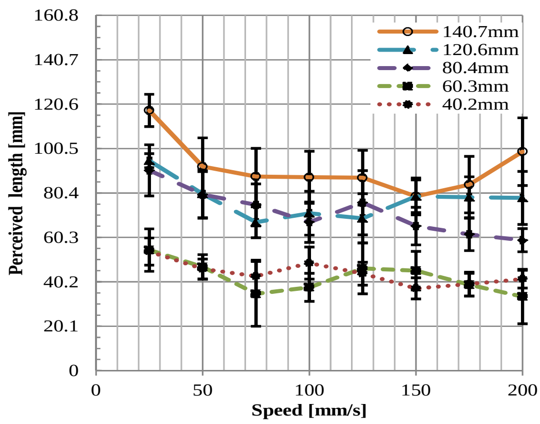
For the 2.5 mm period experiment, we got

$$l_{est\_perc} = 6.7\, v^{-0.11}\, l_{act}^{0.62}. \qquad (4)$$

Since the coefficients of determination for each of the above simple two models accounts for 99%, the models were considered to be adequate.

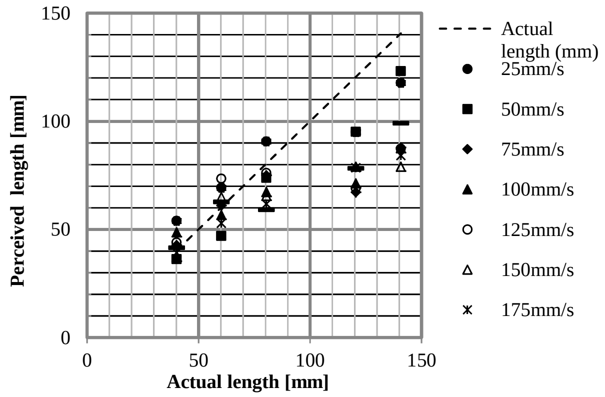
The length-related exponent γ of 0.95 for the 20.1 mm period was very close to 1, and was much superior to that of 0.62 for the 2.5 mm period.

On the other hand, the speed-related exponent β of -0.20 for the 20.1 mm period was a little different from the ideal value of 0, and was inferior to that of -0.11 for the 2.5 mm period. It was considered that, even for the 20.1 mm period, the subjects were no longer able to count dot number in the increased speeds, and suffered larger speed-induced contraction effect. Therefore, it was recommended to employ enough low speeds when using the 20.1 mm period.



(a) Means of perceived lengths (symbols), together with their standard errors(error bars): actual lengths were taken as a parameter.

 ISBN: 978-1-61208-250-9



(b) Means of perceived lengths: speed was taken as a parameter.

Figure 6. Experimental results with the 2.5 mm period raised dots

B. Discussion (Comparison to Existing Works)

In this section, the above explained perceptual models with raised dots were compared with the existing two haptic length perceptual models.

One was a Terada model [3] that was obtained under the cutaneous-sensation-only framework as same as this work, that is, the Terada model was obtained by a cutaneous only sensation for flat sliding surfaces. The model was similar to the Najib model [5]. In Figure 7, the Terada-modeled perceived-lengths are shown by lines, and the 2.5mm period experimental values in this work were shown by symbols. It was interesting that, as a whole, this work of 2.5mm period was good conformity with Terada model except for the data with 140.7 mm length. It suggested that the dot period of 2.5 mm no longer worked as dot planes but as simple planes practically.

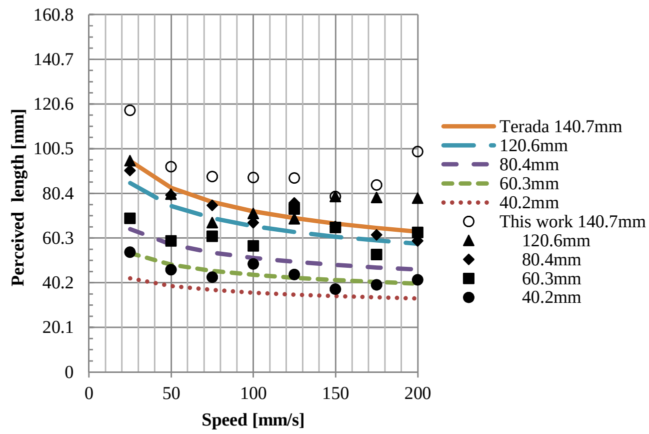


Figure 7. Comparison of the 2.5 mm period experiments to Terada work: the symbols show the means in this work, and the lines show the calculated values by the Terada model.

The other was Hollins model [1] that was obtained under an advantageous framework employing both the cutaneous and proprioceptive sensations. In Figure 8, the Hollins-modeled perceived-lengths are shown by lines, and the 20.1mm period experimental values in this work are shown the symbols. For both the length and speed variations, the 20.1mm period experimental values well agreed to the Hollins model. In Hollins' experiment, not only the cutaneous sensation but also the proprioceptive one being effective in haptic length perceptions was employed. Contrary to the advantageous Hollins framework employing both the cutaneous and proprioceptive sensations, this work of the cutaneous-sensation-only framework might inherently have poor sensitivity. Nevertheless the dot period of 20.1 mm worked as well as the advantageous Hollins model as shown in Figure 7. For reference, the cutaneous sensation was reported to be not much effective by Wouter et al. [5]

Finally, relating to the perception of moving dots, there were some significant works using Braille interfaces, e.g., [10~12]. Using 2×4 dot Braile cells (6.42mm in width, 16.7mm in height), Tahir et al. [11, 12] conducted some experiments on dot pattern selection tasks where 8 kinds of dot patterns were presented either by impulsive way or by 3 Hz of vibration. They obtained an elapsed time of 823ms in average for the vibrating patterns that was a little bit longer than that of 642ms for the impulsive patterns: the comparative merit of impulse to vibration had been agreed with the former work by Pietrzak et al. based on 4×4 dot Braile cells[10]. Here, it would be remarked that the 3 Hz vibration chosen in [11, 12] and the 1/0.6 Hz vibration in [10] respectively corresponded to a speed of 60.3mm/s and 33.5mm/s in the case of 20.1 mm period specimen in this work. The correspondence supported our results that the subjects showed the best performance at the 25 mm/s speed condition in dot number counting task and subject reports. That is, subjects reported that they were no longer able to count the dots when the speed was increased to 50 mm/s and over, and that have changed their perceptual strategy from counting strategy into another cutaneous sensation of depressed point movement perception as in the 2.5mm period dot spacing.

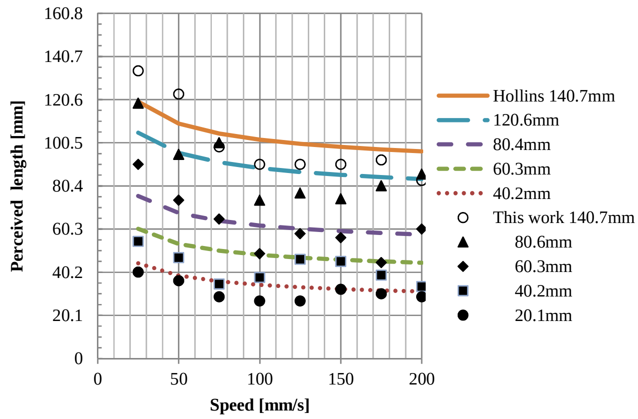


Figure 8. Comparison of 20.1 mm period dot experiments to Hollins work: the symbols show the means in this work, and the lines show the calculated values by the Hollins model.

 ISBN: 978-1-61208-250-9

IV. Conclusion and Future work

Paying notice to the sliding raised dot pathway perception in the passive touch framework, the authors have studied the cutaneous sensation characteristics on index fingerpad.

- By using a power function of actual length l_{act} and velocity v, we formulated a model for the perceived length l_{est_perc} for each of the 20.1 mm and 2.5 mm period dots.
- The length-related exponent γ of 0.95 for the 20.1 mm period was very close to the ideal value of 1, and is much superior to that of 0.62 for the 2.5 mm period. On the other hand, the speed-related exponent β of -0.20 for the 20.1 mm period was a little different from the ideal value of 0, and was inferior to that of -0.11 for the 2.5 mm period.
- In comparisons from relating works, under the cutaneous-sensation-only framework, this work of the 20.1mm period dot showed such a high length perceptual performance as another advantageous framework employing both the cutaneous and proprioceptive sensations. Meanwhile, this work of the 2.5mm period dot showed almost similar performance as with other cutaneous-sensation-only works based on simple planes.

In the future, the authors will make a profound study on the raised-dot-based line perceptual characteristics: increasing sample number, expanding variations of dot period, and extending from one-dimensional lines to two dimensional lines. Furthermore, by using raised dots, they will develop computer-human interfaces, i.e., fingerpad-based line displays.

Acknowledgment

This work was supported by KAKENHI (Grant-in-Aid for Scientific Research (B), No. 21300307) from Japan Society for the Promotion of Science (JSPS)

References

[1] M. Hollins and A. K. Goble, “Perception of the length of voluntary movements,“Somatosensory & Motor Research, Vol. 5, No. 4, pp.335-348 (1988)

[2] L. Armstrong and L. E. Marks, “Haptic perception of linear extent,” Perception & Psychophysics, Vol.61 (6), pp.1211-1226 (1999)

[3] K. Terada, A. Kumazaki, D. Miyata, and A. Ito, “Haptic length display based on cutaneous-proprioceptive integration,” Journal of Robotics and Mechatronics, 18(4), pp.489-498 (2006).

[4] W. M. B. Tiest, L. M. A. Hoff, and A. M. L. Kappers, “Cutaneous and kinaesthetic perception of traversed distance,” Proceedings of the World Haptics Conference (IEEE), pp.593-597(2011)

[5] S. M. N. S. Yusoh, Y. Nomura, R. Sakamoto, and K. Iwabu, “A study on line length and direction perception via cutaneous sensation,” Proc. of AROB 2012

[6] S. J. Lederman, "Skin and Touch," Editor-in-chief, Renato Dulbecco. Academic Press, Vol.7, pp.51-63 (1991)

[7] Y. Wang, M. J. Turner, and W. T. Hewitt, "Creating a novel multiple-frequency based vibration built upon human perception analysis," The 41st United Kingdom Group Meeting on Human Responses to Vibration (2006)

[8] G. Deco, L. Scarano, and S. Soto-Faraco, "Weber’s law in decision making: iIntegrating behavioral data in humans with a neurophysiological model, " The Journal of Neuroscience, 27(42), pp.11192–11200 (2007)

[9] H. Pongrac, "Vibrotactile perception: examining the coding of vibrations and the just noticeable difference under various conditions," Multimedia Systems, Vol.13:pp.297–307 (2008)

[10] T. Pietrzak, I. Pecci, and B. Martin, “Static and dynamic tactile directional cues experiments with VTPlayer mouse”, Eruohaptics 06, pp.63-68 (2006)

[11] M. Tahir, G. Bailly, and E. Lecolinet, "Exploring the impulsion and vibration effects of tactile patterns," British HCI, pp.237-240(2008)

[12] M. Tahir, G. Bailly, E. Lecolinet, and G. Mouret. "TactiMote: Tactile Remote Control for Navigating in Long Lists", ACM International Conference on Multimodal Interaction, pp.285-288 (2008)

1 DOF Tabletop Haptic Mouse for Shape Recognition of 3D Virtual Objects

Hiroshi Suzuki, Hiroaki Yano, Hiroo Iwata
Department of Intelligent Interaction Technologies
University of Tsukuba
Tsukuba, Japan
{suzuki, yano, iwata}@vrlab.esys.tsukuba.ac.jp

***Abstract*—In this paper, we propose a 1 Degree-Of-Freedom (DOF) mouse- shaped haptic device for shape perception of 3D virtual objects. Decrement of DOF of haptic device brings advantages, such as a miniaturization, solidity, and a weight reduction. A 1 DOF haptic device that consists of a built-in optical shaft encoder, two position sensors and a motor, has been developed. This device is easier to integrate with a tabletop system compared to multiple DOF haptic devices. We propose some haptic algorithms, which are effective to 1 DOF haptic device, and two types of pointing environments with a multi-touch overlay. Some experiments were conducted to evaluate the effectiveness of the proposed system. The elements required to make the system functional were clarified.**

Keywords-1 DOF;Mouse Device; Haptic; Image Display; Direct Pointing; Multi-touch Overlay

I. INTRODUCTION

Many haptic devices have been used in the field of 3D shape modeling. They make it possible for users to design virtual characters with intricate shapes, and so on. However, because most haptic devices have multiple DOFs, they tend to be bulky, heavy and expensive. Also, their control algorithms become complex. These factors obstruct the use of haptic devices in general. In terms of the operation of a GUI, we believe that adding 'force feedback' to a tabletop display system is an effective way of further improving its usability and of enhancing its own powers of expression. However, it is difficult to integrate multiple DOF devices with a tabletop display interface because the hardware that is required is quite bulky. We considered that a reduction in the number of degrees of freedom of the haptic interface could be one solution to overcome this issue, since the further the number of degrees of freedom of the haptic device can be reduced, the simpler the hardware that is required becomes.

In 3D shape modeling, we initially not only observe, but also touch the surface of a virtual object to recognize its shape precisely. If a user touches a virtual object on a tabletop display, the user will expect to feel the unevenness of the object. In this process, we assume that there is a principal force vector that is incorporated in the shape recognition process. In order to feel information concerning the unevenness of the surface, depth information regarding the virtual object is a key element in shape recognition. Therefore the 'up-and-down' direction force vector is chosen to present depth information on a tabletop display in this study. As shown in Figure 1, a user can get a sense of touching a virtual object by raising or lowering their fingertip according to the unevenness of the virtual object with visual feedback.

In this research, we developed a prototype system that consists of a 1 DOF mouse-like haptic device and a tabletop display with a multi-touch overlay sensor. The haptic device is small and safe to use. It has the ability to present an 'up-and-down' force to the user's fingertip. It can be controlled in a stable manner by using a simple control algorithm. In addition, we realized a "What You See is What You Touch" (WYSWYT) environment as the pointing environment. In this system, the user can see virtual objects and can also touch the virtual objects directly with his/her fingertip. A number of experiments were conducted to demonstrate the effectiveness of the system.

The remaining of the paper is organized as follows. Section II describes related work. Section III describes the design principle of a haptic interface. Section IV describes the system configuration of our prototype system and its haptic rendering algorithm. Section V describes an evaluation experiment conducted and its results. Finally, the results are discussed in Section VI before concluding.

II. RELATED WORK

Akamatsu reported that providing the sensation of touching a virtual object by imparting vibration to a user's index finger on a mouse interface in a GUI environment is effective for reducing the completion time of a pointing task [1]. Fukunaka proposed a method of presenting a sense of resistance or confliction by using a mouse-shaped device containing a magnetorheological fluid (MR fluid) [2]. However, these devices cannot present the reaction force from the virtual object. They can only present a sense of vibration or collision.

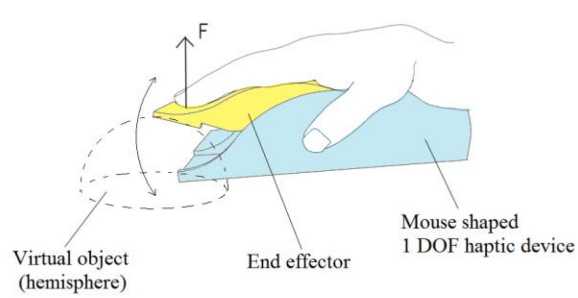


Figure 1. Virtual object presentation by haptic mouse

Marquardt reported how adding a haptic layer to the interactive surface experience can augment the existing visual and touch modalities by using a “haptic tabletop pack”, which can present vertical relief, malleability of materials, and horizontal friction on the surface [3]. But it cannot present shape of the virtual object.

Howe et al. developed the first motion compensated actuated catheter system that enables haptic perception of fast moving tissue structures for surgery of the beating heart. This system enables physicians to operate on the beating heart without caring about deformation of the heart caused by the heart beating, but this system does not enable the user to perceive its shape[4].

The "Touch the Untouchable" system has been developed as an example of a 1 DOF haptic device. This device measures the distance to a remote object by using a laser range sensor, and presents the shape of the object by using the distance data [5]. Alternatively, the "Beyond" system can present virtual objects beyond the display by using a pen-shaped device [6].

The above studies did not reveal a haptic rendering method for complex shaped virtual objects, or consider the influence of reducing the number of degrees of freedom of the interface device. In this study, a prototype 1 DOF haptic mouse has been developed that is suitable for use on a tabletop display. By using this mouse, the effectiveness and limitations of this method were evaluated through a number of experiments.

III. DESIGN PRINCIPLE

In this study, even though the number of degrees of freedom of the haptic device were reduced, we assumed that the ‘visual dominance’ effect imparted by the users could assist the user's shape recognition process. Therefore, as a visual display system, a tabletop GUI environment was selected.

Since the tabletop environment is a flat surface display, the motion of a haptic device on the tabletop is limited by the surface. By combining the limited horizontal motion and depth (up-and-down) motion of the haptic device, the user can recognize the shape of a virtual object stably.

Tabletop haptic interface requires small size, high accuracy, high speed measurement of a user's fingertip position, and sufficient output torque. Furthermore, we assumed to use this device on a tabletop display with visual feedback. Therefore its movable area should be greather or equal to the area of the display area

In order to accomplish this specification, we designed a prototype un-restraining device with built-in sensors and an actuator to securing large movable region. Considering to use the device on a display, the screen of the display should not be coverd over by the device. Hence, a pen shaped device and mouse shaped device can be considered as a structure. However, it is difficult to build in an actuator which can generate sufficent torque on the pen shaped device. Therefore the mouse shaped device, which can store various modules underneath the user's palm, was used in this study.
In order to realize a simple and robust mechanism, the haptic device consists of an end effector, a position sensor, a control PC, and a visual display. The end effector is a lever that can be rotated around a horizontal axis by a DC servomotor so that it can generate an up-and-down force directly to the user's index fingertip. The range of up-and-down motion of the end effector is about 60 mm, taking into account the range of movement of the index finger. The end effector and its base should be transparent to avoid hiding the visual image on the display. An accelerometer, a camera with image processing, an optical mouse sensor are candidates for the position detection sensor. Since the accelerometer has undesireable drift characteristics, and the camera is bulky, two optical mouse sensors are used in this study. They can be measure the position and orientation of the mouse device in un-restraining manner. Also they are cheap and has easy to operate. Moreover the position of the fingertip on the display should be detected without hiding images under the end effector. Therefore, unlike a conventional mouse, neither a position sensor for the mouse nor a stylus could be placed under the end effector. Hence, it is necessary to detect both position and orientation when using this system to calculate the fingertip position. Therefore, 2 sensors are required to detect both position and orientation. However, since the optical sensors can detect only the relative position change of the device, the precise absolute position of the mouse on the display is unacquirable. The direct pointing environment, which is one of the purposes of this research, is difficult by using the sensors. A method to measure the absolute position by using an infrared multi-touch overlay is mentioned later.

IV. 1 DOF HAPTIC MOUSE

A. *System configration*

Figure 2 shows an overview of our prototype for a 1 DOF haptic mouse system. The system consists of a haptic mouse unit, a position detection unit, a control unit, and a PC (OS Windows7 Professional 64bit CPU IntelCorei5 650).

The haptic mouse unit is composed of just a geared(gear ratio 5:1) DC servomotor (*RE25*, made by maxon Inc) and an end effector. Since the gear ratio is small enough, the user can move an index finger up-and-down without feeling resistance. The end effector makes contact with the tip of the user’s index finger. It rotates around a horizontal axis so that it moves the user’s index finger vertically, as shown in Figure 1. The height of the fingertip is calculated from the value of the optical shaft encoder in the motor. The maximum force on the fingertip is 8 N. Two optical mouse

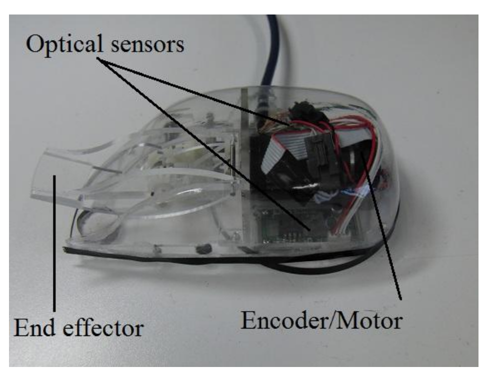


Figure 2. 1 DOF haptic mouse

sensors (made by PixArt Imaging Inc) were used to maintain operability as far as possible. By obtaining the values from the sensors and taking their difference, the relative position and orientation of the device can be calculated. There is an accumulation error of about 4% in terms of the migration distance.

As the control unit, the "Device Art Toolkit" [7] was used to control the optical mouse sensors and the motor of the mouse. This is a controller in which the user can freely change the number and the function of the input-output channels by plugging in various module boards. In this research, a serial communication module, an encoder counter, a PWM output for controlling the motor's output torque, and a motor driver were used.

B. *Haptic Rendering Algorithm*

1) *Penalty method*

In this research, we presented a reaction force in the vertical direction based on the 'penalty method'. As shown in Figure 3, the reaction force was calculated by adding a reaction force and an impulsive force. The reaction force is proportional to the penetration depth (difference between the height of a virtual object and the height of the fingertip position). The impulsive force is proportional to the penetration speed to the object (Equation 1);

$$F = K_p(Z_{obj} - Z_{fin}) + K_d(Z_{fin} - Z_{fin\_p}) \quad (1)$$

where Z_{obj} means the height of the virtual object at the current fingertip position, Z_{fin} means the height of the current fingertip in the case where $(Z_{obj} - Z_{fin}) > 0$: $K_p > 0$. In other cases, $K_p = 0$, Z_{fin_p} means the height of the fingertip in one program loop, in the case of $(Z_{fin} - Z_{fin\_p}) > 0$ and $(Z_{obj} - Z_{fin}) < 0$:$K_d < 0$, in other cases: $K_d = 0$.

A typical example of the locus during touching of a virtual object using the haptic mouse is shown in Figure 4. The green line indicates the locus of the user's fingertip during operation. You can see that the fingertip was moving on the object's surface.

2) *Haptic rendering technique using height map data*

This system can render a virtual object with haptic sensation by using data from a height map, as opposed to the fundamental haptic rendering method using virtual objects defined by some equations or polygons. Since this system renders a virtual object by using only 1 DOF for the upward reaction force, it is possible to render an object by using the height information of the object corresponding to the two-dimensional coordinates on the display surface. We achieved haptic presentation using a height map by utilizing this technique.

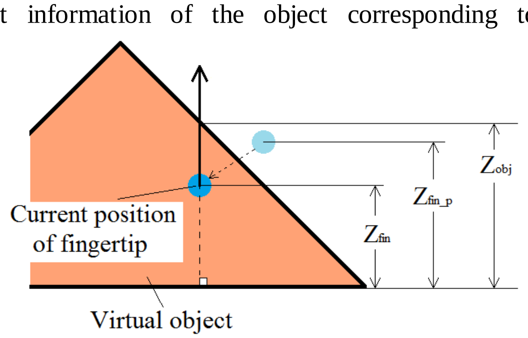


Figure 3. Concept of the penalty method

A height map records height information as brightness information at each pixel of the image. If the color of a pixel is close to black, the height is low. If the color is close to white, the height is high. This system can present a reaction force by using the penalty method described in the previous section by analyzing 24bit bitmap image data. It is possible to render a complex shape of a virtual object by using this technique, even if the shape is difficult to define using equations or a polygon model, as shown in Figure 6.

3) *Haptic rendering method of a virtual wall that is perpendicular to the ground*

When rendering a shape like a cube, which has side surfaces perpendicular to the ground, the user's fingertip is forced to move rapidly upwards. Sometimes this may cause an unwanted vibration and make the user feel uncomfortable. In this study, some small steps were placed on the surfaces to overcome this incongruity. After trial and error, the size of the small steps was set to 1 mm in width and 20 mm in height to reduce the unwanted vibration. The user's fingertip climbs up two steps in a very short time if the virtual wall has a height of 40 mm.

Figure 6 shows a typical locus of a user's fingertip with the haptic mouse when he touches a virtual cube from left to right. By using the proposed technique, the heights of the overshoots are reduced compared with a method that does not include the small steps described in previous section. When the fingertip moves from the upper surface of the cube to the ground, the end effector falls slowly according to the input force at the fingertip. This causes the gently-sloping trajectories on the right side of the cube that are shown in Figure 6.

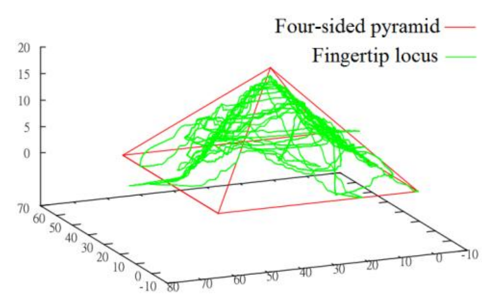


Figure 4. Typical locus of fingertip during touching a virtual pyramid

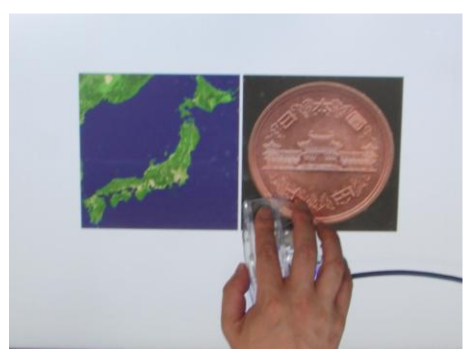


Figure 5. Haptic rendering with a height map data

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-250-9

V. 1 DOF HAPTIC MOUSE WITH MULTI-TOUCH OVERLAY

When using the developed 1 DOF haptic mouse, the spatial relationship between the visual presentation and the haptic presentation was in agreement, and it was thought that presentation closer to the appearance of a real object would be realizable by building a direct pointing environment in which a haptic presentation is shown when touching a visual representation with direct feedback. In this research, in order to harness the characteristics of a 1 DOF haptic mouse that is not grounded, a direct pointing environment was created by combining it with an infrared multi-touch overlay that can detect the positions of two points in an unrestrained manner (Figure 7).

A. Adding a multi-touch overlay for precise measuring

A multi-touch overlay (made by PQLabs) was used in this section. Although the position of an object on the display could be detected accurately by using the multi-touch overlay, there was a time delay of approximately 0.15 s. Since the response of the multi-touch overlay with the infrared sensor was insufficient, we added positional data from the optical mouse sensors during the preceding 0.15 s to the positional data from the multi-touch overlay. In this way, the sensor can avoid an accumulation of errors from the optical sensor, and real-time precise measurements could be realized on the tabletop system. Figure 8 shows a typical measured trajectory of the mouse by using the proposed method at a velocity of 100 mm/s, which is the average velocity when virtual objects are touched by users. The measured data is almost the same as the true trajectory. However, if the velocity is increased to 250 mm/s, approximately 20 mm of overshoot is observed when the mouse decelerates to a halt (Figure 9) due to the delay of the multi-touch overlay. In normal usage, this does not become a concern, since the user does not touch a virtual object at such a high speed or does not stop the mouse suddenly.

The system can measure the position of the fingertip with high precision in real time. The frequency of the measurement was about 1 kHz.

B. Comparison of direct pointing with indirect pointing

An experiment was conducted to evaluate the effects of differences in the pointing environment when using the developed system. In a feasibility study, some users reported that they felt the sense of the presence of the displayed virtual objects with the direct pointing environment compared to the indirect one.

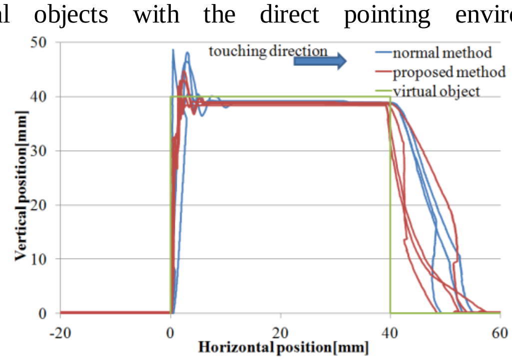


Figure 6. Typical locus of a user's fingertip during touching a virtual cube

Therefore we aimed to reveal the key issues of the different feeling in this experiment.

1) Experimental details

Subjects were asked to perform the task of tracing round a virtual Torus-like object, and their EMG data were recorded, the operation time, the pressure on their fingertip, and the operating locus. In this experiment, 'direct pointing' means a state in which the pointing position on the virtual plane and the position of the subject's fingertip corresponded without displaying the pointer, while 'indirect pointing' means a state in which the position did not corresponded with the displayed pointer (Figure 10). EMG data were measured using an electromyograph (*Active Two*, made by Biosemi) fitted around the extensor digitorum muscle and the flexor digitorum superficialis muscle, which both contribute to the operation of the index finger. A pressure sensor (*FlexiForce*, made by TECKSCAN) was stuck onto the end-effector, and the pressure on the fingertip was measured. Six male subjects conducted 3 trials each under each of these conditions. The order of presentation was randomized in order to negate the influence of the order of the tests.

2) Results

The average of all subjects' operating time is 11.891 seconds in the indirect pointing and 10.745 seconds in the

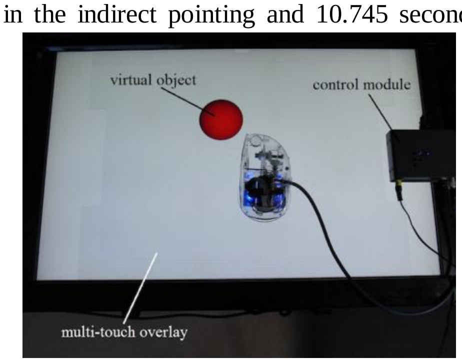


Figure 7. System overview of tabletop haptic interface

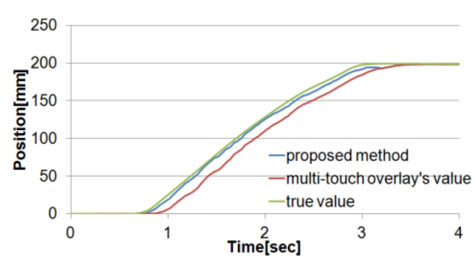


Figure 8. Position detection with two kinds of sensors (100mm/sec)

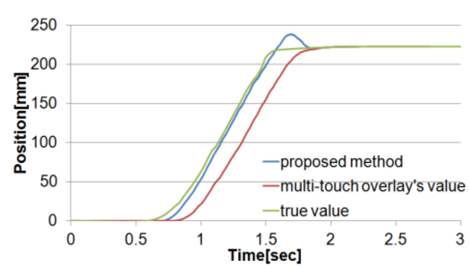


Figure 9. Position detection with two kinds of sensors (250mm/sec)

 ISBN: 978-1-61208-250-9

direct pointing. Operating time was slightly shortened in the direct pointing but there was no significant difference. The average error between each position of a fingertip and target position (peak of Torus) was 2.017 mm in the indirect pointing and 2.112 mm in the direct pointing respectively. Although there was slightly difference of error value between each pointing environment, there was no significant difference.

The root mean square (RMS) values of the EMG measurements were determined in order to perform a waveform comparison. The resulting data was smoothed using a moving-average interval of 200. Moreover, in order to negate the differences between subjects, the peak value of the EMG data for each muscle in the indirect pointing environment was made equal to 100%, the EMG was normalized, and was also normalized with respect to operating time. Next, a conventional 't test' was performed to investigate the difference in EMG for the different types of pointing environments.

Figure 11 shows the average time series of the EMG of the flexor digitorum superficialis muscle obtained as a result of the experiment. The values of EMG, for indirect pointing were stronger in the flexor digitorum superficialis muscle ($t = 15.7969$, $df = 17987$ $p < 0.01$). This means the subjects tended to press the object firmly in the indirect pointing environment. Also, we noted a significant difference in pressure ($t = 17.1802$, $df = 17987$ $p < 0.01$); stronger pressure occurred at the fingertip for the indirect pointing environment. This result is consistent with the EMG data.

In this experiment, the subjects should pay attention to move the mouse pointer on the center of the arc of the virtual circular ring. The subjects tended to place the pointer at the center by using visual information usually. In addition, haptic feedback gave additional information to the subjects. The reaction force rapidly decreased out of the center. If the subject pressed the ring harder, he/she could discriminate the position change easy.

In the indirect pointing environment, since the subjects watched the display and couldn't watch their own hand, which held the mouse, they tended to use not only the visual information but also the haptic information. On the other hand, in the direct pointing environment, the subjects were easy to adjust the position of the mouse by watching their own hand.

These caused the results that the pressure and the EMG in the indirect environment were larger than these in the direct one. In addition, in the direct pointing environment, the subjects tended to press the ring lower pressure at the left or right edge (time of 25% or 75%). Because the ring extends up-and-down direction at these points, the subjects could adjust the position of the fingertip easy.

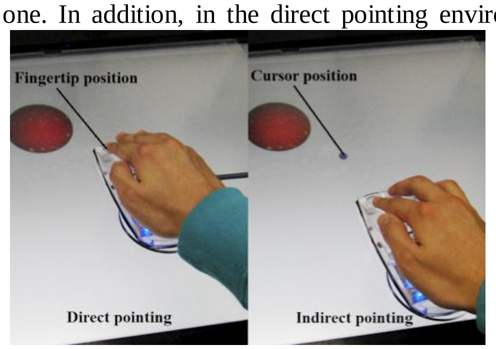


Figure 10. Experimental environment

Moreover, when the subjects were asked to give their impressions after the experiment, a variety of opinions were expressed as to which environment the operativity best supported; the overall opinion was that both environments were almost the same.

VI. DISCUSSION

When compared with an indirect pointing environment, it turned out that the EMG generated in the flexor digitorum superficialis muscle and the pressure on the fingertip in the direct pointing environment were lower. Since the portion of the finger that was directly touching the object in this situation was hidden like it is in the real world, it was believed to influence the subject to touch more carefully.

In the operation of a two-dimensional vision display, it has been reported that, when making a comparison between an indirect pointing environment using a pointing device and a direct pointing environment in which the target is touched with a finger, the direct pointing is superior in terms of operating time and operability [8]. However, there is no significant difference in terms of operating time or the locus of operation in this study. A direct pointing environment has predominance when operating using vision alone, but when haptic presentation is also involved, it can be said that differences in the pointing environments do not have a big influence on the performance of haptic presentation. The direct pointing environment has advantages. It enables a user to touch an object by using his/her own fingertip similar to the real world. Also the user can perceive the shape of a virtual object by combining the device and a real object such as a ruler and so on. However, since implementation of direct

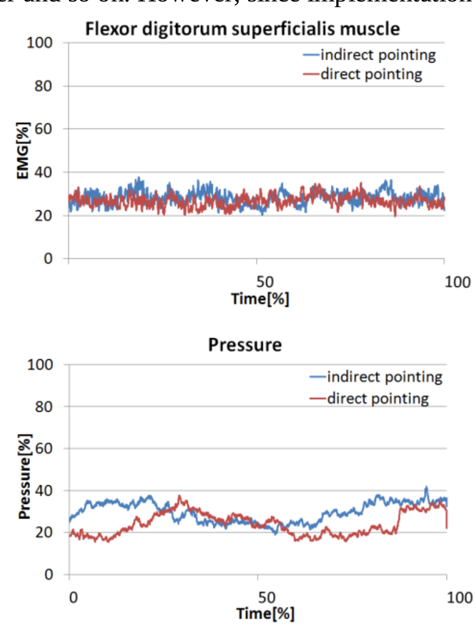


Figure 11. Time series valuation of the average EMG of the flexor digitorum superficialis muscle (top) and average pressure on the end effector of the haptic mouse (bottom)

pointing environment is required capabilities of high accuracy and high-speed position measurement, it is difficult to develop such environment usually. From the previous study, direct pointing visual feedback environment shows superior task performance than that of indirect pointing environment. However, our result means same task performances of the direct pointing environment can be realized with the indirect environment by adding haptic sensation.

When developing a lower DOF haptic device, the haptic rendering algorithm for compensating the decrement of DOF is required. Because 1 DOF mouse , which developed in this research, has only up-and-down haptic presentation function, some techniques are required. For example, as mentioned in this study, a gradually force increasing method to the user's finger to avoid unwanted vibration when presenting a wall of a large cube is required.

Moreover, when a user traces the surface of a quadrangular pyramid with the multiple DOF haptic device, the user can perceive the edge of the pyramid, because the reaction force vector is rapidly change on the edge. However, with 1 DOF device, it has ability to present the up-and-down direction only. When the user's finger overcomes an edge, the reaction force is not change hardly, since the device cannot present horizontal force vector. The user tend to feel a blur edge. In order to solve this, when a reaction force vector changes, increasing method of the feeling of an edge, such as adding vibration is required.

When building a direct pointing environment, it is necessary to implement high-speed and highly precise position detection for optimum fingertip operation. Since the infrared multi-touch overlay used in this research was insufficient in respect of speed, this issue was solved by using a set of optical sensors collectively. To build a direct pointing environment, it would be necessary to solve this problem, possibly by using multiple sensors.

As a result of the experiment, users can trace target positions in high accuracy in both pointing environments. As an application of the proposed system, the operativity of a GUI can be improved. In addition, our system can be applied to a computer aided surgery (CAS) and a surgical training system because the system has capability to present various shapes and elasticity of virtual organs.

There was a report of incongruity about equipment. Since the end effector is a plane attached to the linkage. It is inclined according to the finger's position even though the inclination of a touched surface is not changing. To solve this issue, it should be fabricated with a thimble and a gimbal which center is identical to the center of the fingertip. We plan to change the mechanism of the current end effector.

VII. CONCLUSION AND FUTURE WORK

In this research, we developed a prototype system, consisting of a 1 DOF mouse-like haptic device and a tabletop display with a multi-touch overlay sensor.

By actually constructing such a system including a haptic presentation device featuring a low DOF, the elements required to make the system functional were clarified.

Moreover, it was shown that the pointing environment does not contribute greatly to the performance in haptic presentations by measuring EMG etc.

As a haptic rendering technique, we proposed a virtual wall rendering method that is perpendicular to the ground. As one of other solutions is a method of using the 'visual dominance' in haptic presentation. We think it is possible to touch the vertical wall freely without sense of incongruity by presenting a steep slope as haptic sensation, and an image of the vertical wall as visual sensation simultaneously. We plan to develop such haptic rendering technique.

We also plan to use 'Pseudo-haptics' [9] with our system as a technique for increasing the capability of haptic expression. If a low DOF device is supplemented by visual information, the flexibility of the system can be extended to use with the pseudo haptics technique.

Moreover, although this current system only receives haptic presentation, we plan to improve the system so that a user can change the shape of virtual objects etc.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 22500103.

REFERENCES

[1] Motoyuki Akamastu, Sigeru Sato, "A multi-modal mouse with tactile and force feedback," International Journal of Human-Computer Studies, Volume 40, Issue 3, pages 443-453, 1994.

[2] Ken'ichi Fukunaka, Masataka Imura, Yoshihiro Yasumoto, Yoshitsugu Manabe, Kunihiro Chihara, "Evalution of Haptic Display for Presentation of Softness by MR-Fluid," Proceedings of the Virtual Reality Society of Japan Annual Conference, Volume 9, pages 1C1-3,2004.

[3] Nicolai Marquardt, Miguel A. Nacenta, James E. Young, Sheelagh Carpendale, Saul Greenberg, Ehud Sharlin, "The haptic tabletop pack: Tactile feedback for interactive tabletops," ITS '09 Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, pages 85-92, 2009.

[4] Samuel B. Kesner, Robert D. Howe, "Discriminating Tissue Stiffness with a Haptic Catheter:Feeling the Inside of the Beating Heart", IEEE-World Haptics Conference 2011, pp.13-18, 2011.

[5] Yuichi Miyamoto, Hiroaki Yano, Hiroo Iwata, "Remotr Haptic Preception System Using a Laser Range Finder," Proceedings of the Virtual Reality Society of Japan Annual Conference ,Vol.014, No.HDC03, p.19-22 ,2009.

[6] Jinha Lee and Hiroshi Ishii, "Beyond, collapsible tools and gestures for computational design," Proc of CHIEA '10,pp. 3931-3933, 2010.

[7] Yu-uki Enzaki, Ryota Sato, Hiroaki Yano, Hiroo Iwata, "Development of Device Art Toolkit," Proceedings of the Virtual Reality Society of Japan Annual Conference, Vol. 15 No. 3, pp.417-426, 2010.

[8] Evan D. Graham, Christine L. MacKenzie, "Physical versus virtual pointing," CHI '96 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 292 – 299, 1996.

[9] Anatole Lécuyer, **"**Simulating Haptic Feedback Using Vision," A Survey of Research and Applications of Pseudo-Haptic Feedback, Precence,18(1)pp.39-53, 2009.

Usability Study of Static/Dynamic Gestures and Haptic Input as Interfaces to 3D Games

Farzin Farhadi-Niaki[1], Jesse Gerroir[2], Ali Arya[3], S. Ali Etemad[4], Robert Laganière[5], Pierre Payeur[6], Robert Biddle[7]

[1] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada, ffarh101@uottawa.ca
[2,3] School of Information Technology, Carleton University, Ottawa, Canada
[4] Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada
[5,6] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada
[7] Department of Computer Science, Carleton University, Ottawa, Canada

***Abstract*— In this paper, the quality of the interaction of users with a 3D game using different modalities is studied. Three different interaction methods with a 3D virtual environment are considered: a haptic 3D mouse, natural static gestures (postures), and natural dynamic (kinetics) gestures. Through a comprehensive user experiment we compared the pre-defined natural gestures to each other and also to a haptic interface which is designed for the same game. The experiments analyze precision (error), efficiency (time), ease-of-use, pleasantness, fatigue, naturalness, mobility, and overall satisfaction as evaluation criteria. We also used user-selected ranks of importance as weight values for evaluation criteria to measure the overall satisfaction. Finally, our user experiment presents a learning curve for each of the three input methods which along with the other findings can be a good source for further research in the field of natural multimodal Human-Computer Interaction.**

Keywords— Usability study, static/dynamic gestures, haptics, 3D game, human factors.

I. INTRODUCTION

Due to the substantial growth in computational capabilities, both in computer hardware and software, human computer interaction (HCI) systems have progressed, becoming more accurate, efficient, and practical.

In general, two major aspects of HCI systems need extensive study and analysis: functionality and usability. The reason for such categorization is that while the former is often bound by machine hardware, software, and complex algorithms, the latter depends largely on another complex process: the human experience.

Recently, different aspects of HCI have been subject to research and improvement. Modalities involved in such systems often play a determining role in the outcome and experience of employing the device. Multimodal interfaces (i.e. those with different input/output methods) are becoming more popular due to their variety, efficiency, and ability to adapt to user needs [1][2]. Active systems, capable of dynamically and intelligently adapting to different scenarios are also becoming more practical. Although there have been some advances in research on multimodal interfaces, numerous questions remain unanswered and until today, very few systems have integrated multiple modalities flawlessly and effectively [3]. Among these questions are:

1) What modalities are more suited to different tasks?

2) What is the right balance of input/output methods in multimodal systems?

The first question is the main focus of this paper where different human factors when playing a 3D computer game are measured, studied and analyzed. Following our previous studies on usability of natural interfaces [1], in this paper we have focused on two input methods that try to provide a more natural interaction: gesture-based input using a 3D camera and haptic force-feedback input using a 3D mouse. For gesture-based method we have considered static hand/finger gestures and dynamic arm gestures. By “static” we refer to those gestures that are defined by a single state. The term “dynamic” is used to refer to gestures that are identified by a certain movement. This provides three separate input options, all applied to a 3D computer game, a simple slingshot game developed using Microsoft XNA. Successive to implementation, the system is tested with multiple users which provide the feedback needed to analyze the usability of such systems with respect to factors such as precision, efficiency, ease-of-use, fun-to-use, fatigue, naturalness, mobility, and overall satisfaction.

The major contributions of this study are: a) choice of natural gestures, b) system design (UI and gesture recognition) and novel use of existing API’s to implement gesture recognition and haptic force feedback methods, c) usability study for gesture-based and haptic 3D mouse inputs, d) overall satisfaction analyses, directly from the users feedback, and indirectly from average of multiplying the normalized weighted satisfaction criteria, and e) presenting a learning curve for each of the three modalities.

In the following sections, the complete process of construction of our proposed system is discussed. In Section 2 a review of some key literature and usability studies in the field of gestural and haptic HCI is carried out. Section 3 deals with methodology including the UI design, natural gestures selection, gesture and haptic recognition modules along with their algorithms designed to control the UI, and

 ISBN: 978-1-61208-250-9

eventually more detail of the experiment process and our evaluation method. In Section 4 the experimental results are analyzed and discussed. Finally in Section 5 some concluding remarks are presented.

II. Related Work

A system, capable of recognizing human gestures and providing haptic feedback, is considered a major step towards a more natural and viable multimodal system. For example, in an augmented reality system to play table tennis, it would be ideal if the user could provide the system with controlling commands using either speech or gesture and feel force feedback when they hit the ball using the racquet. In such systems, human gestures play a critical role which needs to be preferably detected, classified, and interpreted through computer vision and pattern recognition means in order to avoid "non-natural" sensors.

A. Gestures

Humans, and many other living organisms, have been employing motions of limbs for expression. The broad definitions for gestures in biology and sociology allow researchers to describe the gestures proprietarily [4]. Kendon [5] classifies a gesture as: gesticulation (impulsive movements of hands/arms during speech), language-like gestures (replace words/phrases), pantomime-like gestures (depict objects/actions), emblems (common gestures, e.g., the "V" sign for victory), and sign languages (sets of gestures/postures defining linguistic communication systems, e.g., ASL, the American Sign Language). From gesticulation to sign languages, the association with speech decreases and social parameter increases.

When interacting with computers, gestures can be utilized through vision based techniques. Hand movements and poses such as pointing, grabbing, and moving, are extremely intuitive and content rich (both spatiotemporally and perceptually), and therefore, perfect means for interaction purposes [3]. For eventual gesture utilization in an HCI system, modeling (2D, 3D, or 4D), analysis (feature extraction through region of interest), and recognition (combined features to provide the scene's information) need to be designed precisely [6].

Although static arm/hand gestures, also called hand postures (shape-based recognition algorithms), have had the main focus in gesture recognition field, recently researchers are showing more interests in incorporating dynamic gestures (temporal-based classification systems) in their study, due to broader domain of hand's dynamic actions comparing to hand postures [7].

B. Vision-Based Modality

There has been significant study in the field of vision-based modalities. The hands and line of sight (LoS) combination, as the interaction method, can lessen the fatigue compared to a one hand pointing interaction [8].

Examples by Villaroman et al. [9] are presented to demonstrate how Kinect-assisted instruction can be utilized to accomplish certain learning results in Human Computer Interaction (HCI) courses. Moreover, the authors have confirmed that OpenNI, in addition to its accompanying libraries, are adequate and beneficial in enabling Kinect-assisted learning activities.

Based on a usability evaluation, Bhuiyan and Picking [10] recommend that a gesture user interface application, titled Open Gesture (available for standard tasks, such as making telephone calls and operating the television), can improve the lives of the elderly and the disabled users by creating more independence while some challenges still remain to be overcome.

An experimental study by Ebert et al. [11], on touch-free navigation through radiological images, measured the response period and the practicality of the system compared to the mouse/keyboard control. The image recreation time using gestures was 1.4 time longer than using mouse/keyboard. However it does remove the risk of infection, for both patients and staff.

In spite of all developing and improving facts in above mentioned works, it seems there is a significant lack of studies in terms of natural gestures selection. Designing a suitable user interface for the following usability studies is also crucial. Finally, we believe that there are more usability features which need to be studied than those in above mentioned works.

C. Haptic Modality

Haptics denotes the human's sense of touch for feeling or manipulating a virtual object. Haptics has been supported by a collaboration of various disciplines such as psychophysics, neuroscience, biomechanics, robot design, modeling and simulation, and software engineering. A wide range of applications have emerged and span many fields such as product design, education, video games, graphic arts, medical trainers, and rehabilitation [12].

Regarding the main forms of feedback, haptic devices can be classified into three groups [7]: ground referenced force feedback (e.g., Phantom device [13]), body referenced force feedback (e.g., CyberGrasp [14]), and tactile feedback (e.g., CyberTouch [14]).

Considering the continuous contact of user with a haptic device in most haptic systems, the user's perception of the virtual environment is hindered, and it is also impossible for the user to feel a new tactile sensation. The latter drawback hampers the integration of tactile and force feedbacks in one haptic device [15][16]. Moreover, most haptic devices have a very limited workspace (e.g., 13cm × 18cm × 25cm in the Phantom Premium 1.0A model [17][18]). Therefore, constant contact in such a limited space impairs to incorporate rich interaction elements in an extended virtual environment. To overcome these limitations and to generate an inclusive haptic experience, Ye [3] designed an augmented reality system that employs visual tracking to seamlessly merge force feedback with tactile feedback.

To evaluate what can be profoundly achieved in creating synthetic haptic experiences, technology development and quantitative investigations are indispensable. Having this combination, preferably in a multimodal system, would help to assign the right balance of input/output methods to different tasks in order to build an interactive system

 ISBN: 978-1-61208-250-9

between human and machine as much natural and effective as possible.

III. Methodology

We initially defined a series of static (Fist and Palm) and dynamic (Circle and Push) gestures to be linked to the tasks in our test game. Then we designed proper algorithms to detect our predefined arm/finger gestures using the Kinect sensor and relatively novel existing API's (OpenNI [19], NITE [20], OpenCV) and to interact with our slingshot 3D game interface (using Microsoft XNA). Other than the vision-based modality, we incorporated a haptic modality to also interact with the same game interface (Figure 1).

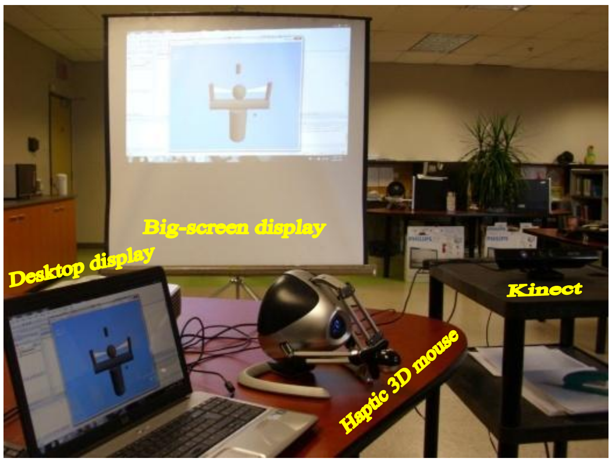


Figure 1. The experiment devices (Falcon haptic 3D mouse, Microsoft Kinect camera, big-screen display using a high quality projector), and testing environment.

Using the static gestures, our prototype grabs the virtual ball when a fist is detected in the ball's space, then releases it when it recognizes a palm. This process is reproduced in dynamic gestures by detecting a circle drawing in the ball's space to grab it, and pushing to throw the ball towards a pin object for scoring purpose. Furthermore, using a 3D Falcon haptic device (the world's first consumer 3D touch device, which allows users to use their sense of touch in computing), the user grabs the ball (similarly to the vision-based procedure, when the 3D position of the pointer is in contact with the ball's space) by pressing the button (at this point the force feedback of the elastic band is felt), and throws the ball by releasing the button.

We then statistically compared the three input methods (static postures, kinetic/dynamic gestures, and haptic force feedback) considering the following human factors: precision (errors), efficiency (time), ease-of-use, fun-to-use, fatigue, naturalness, mobility, and overall satisfaction. As another contribution in this research we analyzed the overall satisfaction of users in two ways: i) directly from the users feedback, and ii) indirectly from average of **multiplying *the normalized weighted satisfaction criteria*** (easiness, pleasantness, fatigue, naturalness, and mobility), which have been ranked separately by users, **to *the respected satisfaction feedback*** rated from users per devices, in order to provide a more significant and reliable rate for a practical overall satisfaction (more details in section IV.A.8.b).

Finally, our user experiment presents a learning curve for each of the three modalities by recording the time between any hit occurrences for the first 10 successful shots since the beginning of training session.

A. *User Interface*

For the user interface we have designed a slingshot 3D game virtual environment (using Microsoft XNA). The design is kept as simple and minimalistic as possible, with neutral colors to reduce user error or bias (Figure 2). An option for changing the camera view is also included that users can use when they wish. The same virtual environment is controlled by the three modalities (static postures, dynamic gestures, and haptic controller) independently.

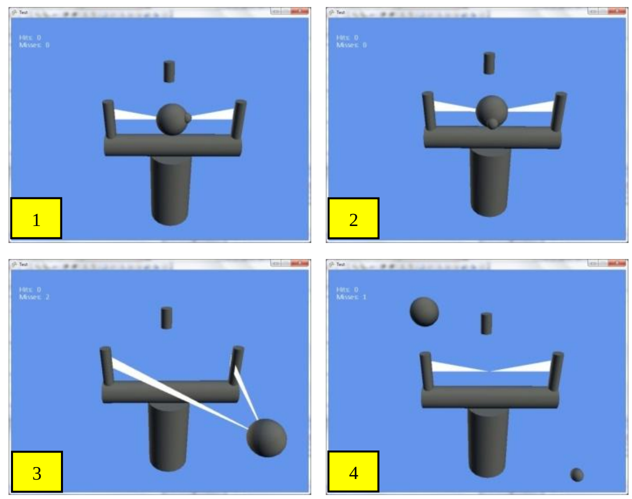


Figure 2. Slingshot 3D game virtual environment.

B. *Vision-Based Module*

Using the Kinect unit (as a commonly used vision-based input device) enables us to identify the depth of every single pixel in the frame by projecting a pattern of dots with the near infrared projection over the scene, and establishing the parallax shift of the dot pattern for each pixel in the detector. In addition, using OpenNI and NITE has helped us to secure our system with a higher stability and efficiency, and to develop a capable algorithm to recognize the arm and finger gestures.

Using the above explained method we can conserve the developing time (no need for making samples and efforts in training/testing sessions) and running time (CPU performance) for gesture recognition and user interaction compared to traditional learning-based method.

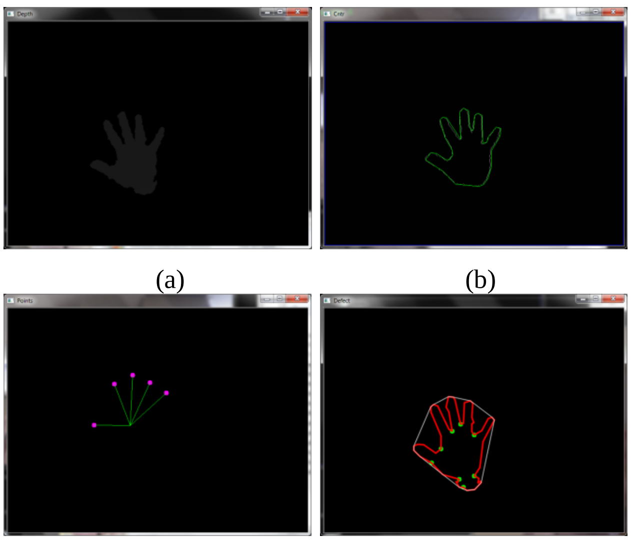
1) *Static Gestures*

Considering the natural gestures to represent the tasks in interaction with our slingshot game interface, the selected static postures definition is presented in Table I.

TABLE I. DESIGN FOR FINGER GESTURES SET.

Description	***Fisting***	***Palming***
Static gesture		
Action	***Grabbing the ball***	***Releasing the ball***

In order to recognize these gestures, as shown in Figure 3, we first needed to detect the fingertips through: i) depth thresholding, ii) contour extraction, and iii) assuming vertices of convex hull to be fingertips if their interior angles of hull corners (T_θ is the angle spanned by the predecessor edge and successor edge incident to vertex facing to the inside of contour polygon) are small enough.



(a) (b) (c) (d)

Figure 3. Images in hand and finger detection processes, (a) depth thresholding, (b) contour extraction, (c) fingertips detection, and (d) convexity defects (depth points between two fingers).

$$\boldsymbol{T_\theta} = \cos^{-1}\frac{(\boldsymbol{v_x} \times \boldsymbol{v'_x}) + (\boldsymbol{v_y} \times \boldsymbol{v'_y})}{\|\boldsymbol{v}\| \times \|\boldsymbol{v'}\|} \qquad (1)$$

where $\boldsymbol{v_{x,y}} = \left|\text{successor of corner index } (\boldsymbol{x}, \boldsymbol{y}) - \text{corner index } (\boldsymbol{x}, \boldsymbol{y})\right|$ and $\boldsymbol{v'_{x,y}} = \left|\text{predecessor of corner index } (\boldsymbol{x}, \boldsymbol{y}) - \text{corner index } (\boldsymbol{x}, \boldsymbol{y})\right|$. Moreover, we cropped the depth map to remove the wrist out of the frame in order to improve accuracy.

Following the above process of fingers detection, the system will recognize fist (no fingers) and palm (at least four fingers), and interact with the user interface through Algorithm 1.

Algorithm 1. The algorithm controlling UI using gesture recognition.

```
if (an initial Wave gesture happens) then
  the pointer appears (session starts)
  while the pointer follows the hand position (session updates)
    if (a grabbing gesture Fist/Circle is detected in the object area) then
      grab the object (save (x,y,z) based on distance deviation)
      if (a releasing gesture Palm/Push is detected in the object area) then
        release the object (transfer data)
```

2) Dynamic Gestures

Based on the available gesture recognition module in NITE, we have selected the following dynamic gestures definition as shown in Table II. During the controlling sessions of OpenNI and built-in gestures in NITE, the system will recognize circle and push, and interact with the user interface through the similar algorithm mentioned for finger gesture recognition (Algorithm 1), only replacing Fist and Palm gestures to Circle and Push.

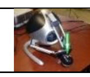
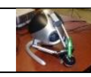
TABLE II. DESIGN FOR ARM GESTURES SET.

Description	***Circling***	***Pushing***
Dynamic gesture		
Action	***Grabbing the ball***	***Releasing the ball***

C. Haptic Module

We designed our haptic events controller traditionally, as shown in Table III. As the user moves the grip in three dimensions (right-left and, up-down like a mouse, but also forwards-backwards, unlike a mouse), the Falcon's software keeps track of where the grip is moved and creates forces that a user can feel. The default grip is a small spherical grip with 4 buttons on the top.

TABLE III. DESIGN FOR HAPTIC CONTROLLER.

Description	***Click & Hold the button***	***Release the button***
Haptic 3D mouse		
Action	***Grabbing the ball***	***Releasing the ball***

D. User Experiments

Evaluation of the different interface modalities will be based on a number of criteria. These criteria are summarized in Table IV. Users have been asked to rate each criterion on a scale of 1 to 7. To ensure an unbiased sequence of used modalities during the user experiments (training and test sessions) we divided the participants into three groups (A, B, C) and permutated them such that each mode (M_1, M_2, M_3) had an equal share of sequence as first, second, and third used modalities (ABC, BCA, CAB).

1) Training Session

The experimental evaluation starts with a training session of about 15 minutes (five minutes on each device) until the participants feel comfortable to start the test session. The participants try the following four primitive tasks to get used to the application in order to run the test session precisely:

a) Move the pointer (tool) around

b) Grab the ball and point towards the pin (object)

c) Pull the ball (elastic band)

d) Throw the ball towards the pin (object)

Finally, the time between any hit occurrences for the first 10 successful shots is recorded since the beginning of training session, in order to study the learning curve.

TABLE IV. EVALUATION CRITERIA AND THEIR REPLYING CONTEXTS.

Evaluation Criteria	**Weighted Coefficients** (ranked by participants)	**Questions** (answered by participants)	**Measurements** (measured by the observer)
Ease of Use	$\tilde{E}$	How easy was it? (E)	Normalized $\tilde{E}$ ($\tilde{E}_N$)
Fatigue	$\tilde{F}$	How non-fatiguing was it? (F)	Normalized $\tilde{F}$ ($\tilde{F}_N$)
Naturalness	$\tilde{N}$	How natural was it? (N)	Normalized $\tilde{N}$ ($\tilde{N}_N$)
Pleasantness	$\tilde{P}$	How pleasant was it? (P)	Normalized $\tilde{P}$ ($\tilde{P}_N$)
Mobility	$\tilde{M}$	How flexible was it? (M)	Normalized $\tilde{M}$ ($\tilde{M}_N$)
Overall Satisfaction		How satisfied are you overall? (S)	Adjusted-Weighted-Overall Satisfaction ($\tilde{S}$)
Efficiency			Required time for 5 shots
Effectiveness			Error rate (misses in 5 shots)
Learning Curve			Time rate for the first 10 hits

2) Test Session

The second part of the evaluation is a test session during which the user tries 5 shots using the haptic 3D mouse, static postures and kinetic gestures to complete a later satisfaction questionnaire. An experiment observer keeps a record of the total time (during the 5 shots) and the scores (number of hits/misses out of 5 shots) from the start to the end point.

IV. RESULTS AND DISCUSSIONS

This study has been conducted using 20 participants (11 males and 9 females). 17 participants were right-handed. They ranged in age from 15 to 45, with an average age of 29 years old. Ethics approval was secured for participant surveys. None of participants were familiar with the use of our three modalities before. The participants first read the experiment instructions and were given introductions to the tasks they were to complete during the trial.

The trial was divided into three phases:

1) *Training phase:* to get familiar with the applications
2) *Test phase:* the main process to observe the results
3) *Satisfaction phase:* to complete a questionnaire

A. Analyses

To analyze the effects of the different human factors being studied, one-way repeated measures analysis of variances (ANOVA) [21] is carried out in MATLAB, for the modality/input device variable (haptics vs. static postures vs. dynamic gestures). All analysis are concluded at $p < 0.05$ significance level and for 20 participants.

Notation: Through the following analyses of human factors, we calculate the mean and standard deviation for different variables in the forms of $M^{\text{human factor}}_{\text{variable}}$ (e.g., $M^{\text{Time}}_{\text{H}}$ is the mean of time for haptic device) and $SD^{\text{human factor}}_{\text{variable}}$ (e.g., $SD^{\text{Precision}}_{\text{D}}$ and $SD^{\text{Precision}}_{\text{S}}$ are the standard deviations of precision for dynamic and static gestures). Moreover, F(df,MS) is the test statistic (F-ratio) in which df and MS are the degree of freedom and mean square respectively for the variables (within subjects). The F-ratio is calculated using MS*variable*/MS*error* and P is the probability value.

1) Time

The analysis illustrates that for time (duration of test session), there is significant effect in the modality, F(2,186.189) = 6.375, P = 0.0041. Further repeated measures of ANOVA were used to make post hoc comparisons between each paired modalities. This reveals that the effect of dynamic gestures vs. static postures is significant on time, F(1,370.88) = 15.76, P = 0.0008, where static postures show to be significantly faster compared to dynamic gestures (Figure 4).

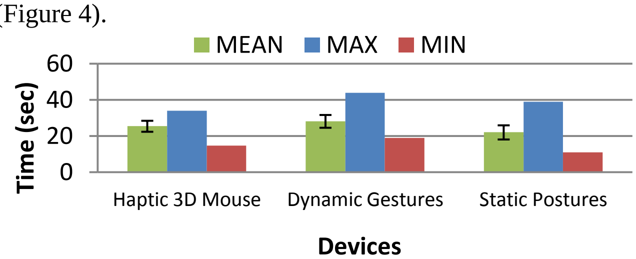


Figure 4. Temporal statistical facts.

2) Precision

The analysis illustrates that the modality has significant effect on the precision (number of hits in five shots) when playing the game, F(2,3.80) = 5.92, P = 0.0058. Paired repeated measure ANOVA shows significant difference on scores between haptic 3D mouse vs. static postures, F(1,4.90) = 7.11, P = 0.015, and between dynamic gestures vs. static postures, F(1,6.40) = 9.65, P = 0.0058, where static postures are significantly more precise than the two other modalities (Figures 5 and 6).

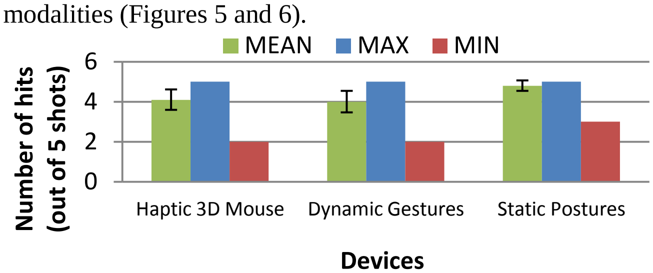


Figure 5. Score-based statistical facts.

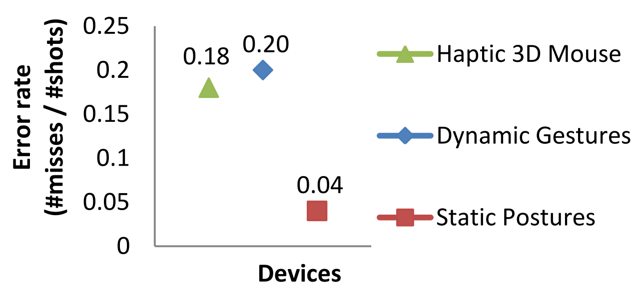


Figure 6. Error rate per device.

3) Easiness

Analyzing the results indicates significant effect for the modality on easiness (how easy to interact with the UI) ratings, F(2,5.55) = 5.42, P = 0.0085. Paired analyses of variances indicate that the haptic 3D mouse is significantly easier to be worked with compared to dynamic gestures, F(1,3.60) = 4.75, P = 0.0421, while static postures are significantly easier than dynamic gestures, F(1,11.02) = 7.91, P = 0.0111 (as shown in Figure 7).

4) Fatigue

The analysis on feedback for fatigue (how fatiguing to interact with the UI) reveals that the modality has significant effect F(2,8.02) = 6.93, P = 0.0027. Paired analyses reveal that the static postures modality is significantly less fatiguing than the haptic 3D mouse, F(1,9.02) = 7.01, P = 0.0159. Furthermore, it is revealed that static postures are significantly less fatiguing than dynamic gestures, F(1,14.40) = 15.54, P = 0.0009 (as shown in Figure 7).

5) Naturalness

For this factor (how natural to interact with the UI), it is shown that the modality has significant effect, F(2,9.62) = 11.63, P = 0.0001. Paired analyses indicate that static postures are significantly more natural than both the haptic 3D mouse F(1,18.22) = 18.95, P = 0.0003, and the dynamics gestures, F(1,9.02) = 16.37, P = 0.0007 (Figure 7).

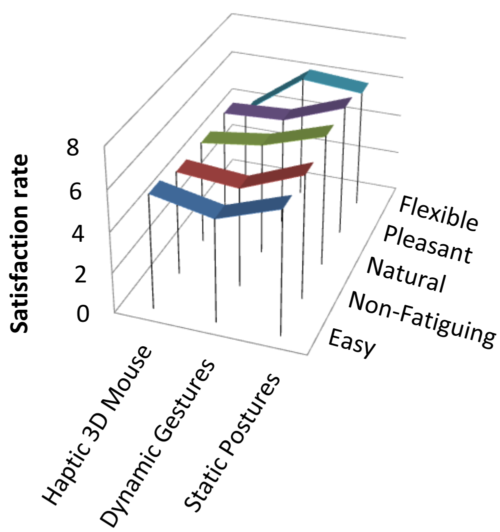


Figure 7. Direct-MEAN: rate of satisfaction criteria (direct criteria feedback).

6) Pleasantness

When analyzing the participants' feedback for pleasantness (how pleasant to interact with the UI), a similar trend to that of naturalness is observed. The effect of modality is significant, F(2,9.32) = 9.15, P = 0.0006. Moreover, similar to naturalness, paired analyses indicate that static postures are significantly more pleasant than both the haptic 3D mouse, F(1,15.62) = 19.96, P = 0.0003, and dynamics gestures, F(1,12.10) = 17.82, P = 0.0005 (as shown in Figure 7).

7) Mobility

Analysis for participants feedback on mobility (the working space to interact with the UI), shows significant effect, F(2,27.65) = 20.72, P = 0.0000. Paired analyses indicate that participants experience significantly less mobility when using the haptic 3D mouse compared to dynamic gestures, F(1,46.22) = 24.212, P = 0.0001, and static postures, F(1,36.10) = 31.32, P = 0.0000 (Figure 7).

8) Overall Satisfaction

a) Directly Gained (S)

In the overall satisfaction (how overall satisfactory to interact with the UI) rating obtained from participants, significant effect is observed, F(2,5.12) = 5.76, P = 0.0065. Paired analyses indicate that static postures are significantly more satisfactory than both the haptic 3D mouse, F(1,7.22) = 9.62, P = 0.0059, and the dynamic gestures, F(1,8.10) = 9.11, P = 0.0071 (as shown in Figure 7).

b) Indirectly Gained ($\tilde{S}$)

As part of the questionnaire, the participants were asked to rank the five satisfaction criteria (easiness, non-fatigue, naturalness, pleasantness, and mobility) from 1 to 5 (the more important satisfaction factor gets the higher rank) to playing a 3D computer game (Figure 8). We name these new parameters as weighted satisfaction criteria or weighted coefficients ($\tilde{E}$, $\tilde{F}$, $\tilde{N}$, $\tilde{P}$, $\tilde{M}$). In order to refine the results and acquire an unbiased set of coefficients, we normalize the rank of satisfaction criteria per user through the following equation:

$$\tilde{X}_N = \frac{\tilde{X}}{\sum_{i=1}^{5} \tilde{X}(i)} \quad (2)$$

where $\tilde{X}_N$ is the normalized rank of satisfaction criteria per user (Figure 9); and $\tilde{X}$ is the weighted coefficient ($\tilde{E}$, $\tilde{F}$, $\tilde{N}$, $\tilde{P}$, $\tilde{M}$), ranked from each user for satisfaction criteria.

We produce the weighted rate of satisfaction criteria (weighted criteria feedback) per device and per user (ω) through the following equation:

$$\omega = X \times \tilde{X}_N \quad (3)$$

where X is the rate of satisfaction criteria (direct criteria feedback) per device and per user (E, F, N, P, M).

Finally, we infer a new practical rate ($\tilde{S}$) defined as Adjusted-Weighted-Overall Satisfaction (average of weighted criteria feedback) per device and per user through the following equation:

$$\tilde{S} = \frac{\sum_{i=1}^{5} \omega(i)}{5} \quad (4)$$

The modality shows significant effect on the indirectly gained (computed) overall satisfaction, F(2,0.29) = 11.54, P = 0.0001. Paired analyses reveal that similar to the overall satisfaction rates acquired directly from participants, static postures are significantly more satisfactory compared to the haptic 3D mouse, F(1,0.54) = 25.95, P = 0.0001, and to the dynamic gestures, F(1,0.30) = 13.48, P = 0.0016 (Figure 10).

c) Comparison

Figure 11 illustrates a comparison between the directly and indirectly gained "overall satisfaction" (S vs. $\tilde{S}$).

 ISBN: 978-1-61208-250-9

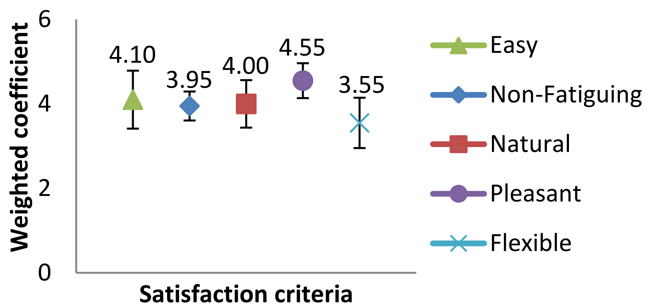


Figure 8. Rank of satisfaction criteria (weighted coefficient).

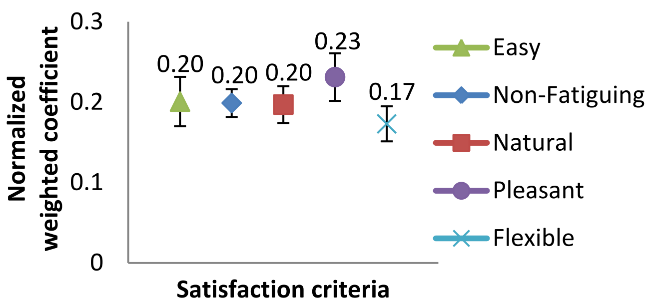


Figure 9. Normalized rank of satisfaction criteria (normalized weighted coefficient).

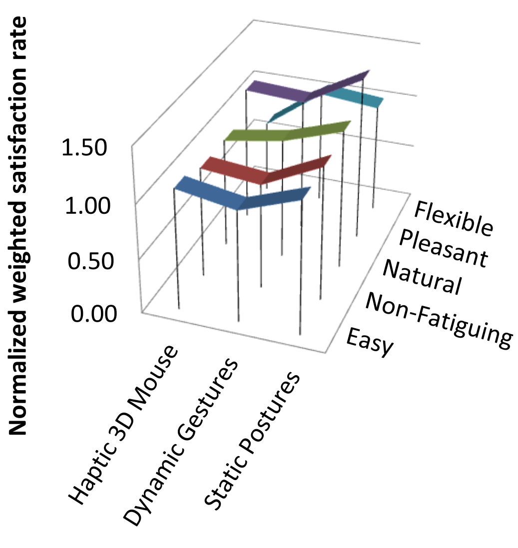


Figure 10. Weighted-MEAN: weighted rate of satisfaction criteria (weighted criteria feedback).

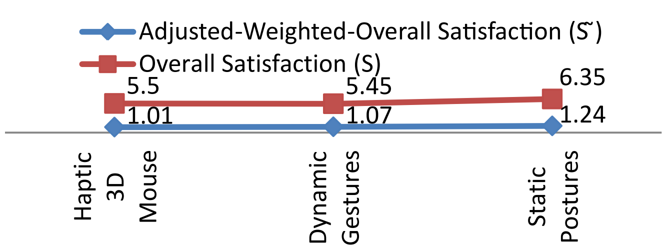


Figure 11. Overall satisfaction (gained directly vs. indirectly).

9) Four Primitive Tasks

Through the experiments, the participants were also asked to rate their satisfaction for the main tasks of moving the pointer, grabbing the ball, moving the ball, and throwing the ball for each device.

For moving the ball, the modality shows significant effect, $F(2,4.62) = 5.14$, $P = 0.0105$. Paired analyses indicate that it is significantly easier to move the pointer in static postures application compared to dynamic gestures application, $F(1,9.02) = 11.081$, $P = 0.0035$.

When grabbing the ball, the effect is significant also, $F(2,13.07) = 17.82$, $P = 0.0000$. Paired analyses show that all three modalities cause significant differences. In other words, it is significantly easier to grab the ball using haptic 3D mouse compared to dynamic gestures, $F(1,10.00) = 14.61$, $P = 0.0011$, and it is significantly easier to grab the ball using static postures compared to the haptic 3D mouse, $F(1,3.60) = 4.44$, $P = 0.0486$, and dynamic gestures, $F(1,25.60) = 36.30$, $P = 0.0000$.

For pulling the ball, the effect is again significant, $F(2,4.82) = 4.61$, $P = 0.0161$. It is significantly easier to pull the ball in static postures application than using haptic 3D mouse, $F(1,7.22) = 11.18$, $P = 0.0034$, and in dynamic gestures application, $F(1,7.22) = 6.16$, $P = 0.0226$.

Finally for throwing the ball, the effect is significant, $F(2,7.72) = 8.65$, $P = 0.0008$. It is significantly more difficult to throw the ball using dynamic gestures than using both haptic 3D mouse, $F(1,11.02) = 9.32$, $P = 0.0065$, and static postures, $F(1,12.10) = 10.50$, $P = 0.0043$.

Figure 12 illustrates the comparisons among the tasks and devices.

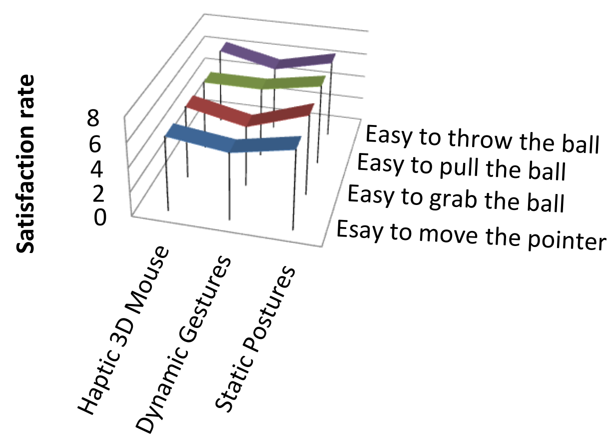


Figure 12. Satisfaction rate for primitive tasks.

10) Learning Curve

In order to study the learning curve, we have recorded the time between any hit occurrences for the first 10 successful shots since the test session starts (Figure 13). This data also presents an initial speed rate (hits/sec) during the beginning of the training session (Figure 14 shows the reversed speed).

For the average learning curves acquired for different modalities, there is significant effect, $F(2,6.38) = 4.36$, $P = 0.0286$. Paired analyses reveal that the mean learning time

 ISBN: 978-1-61208-250-9

per hit for the static posture is significantly less than that of the haptic 3D mouse modality, F(1,12.72) = 5.89, P = 0.0381.

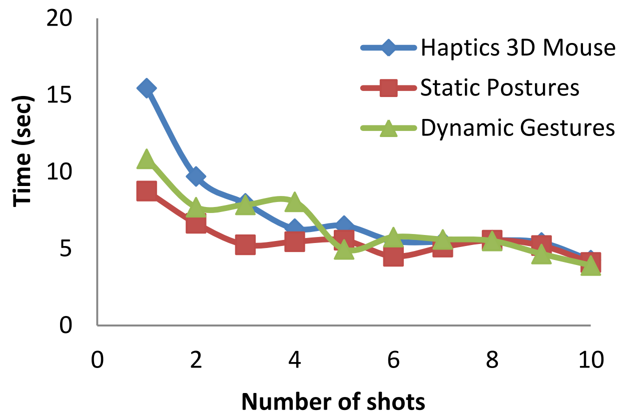


Figure 13. Learning Curves.

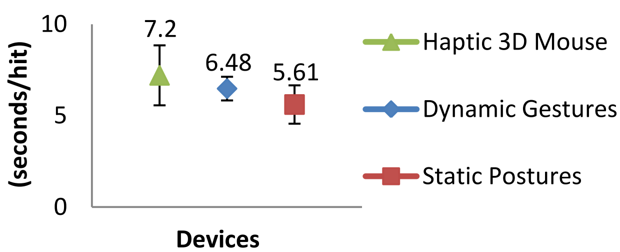


Figure 14. Reverse speed (average time (sec) per hit): MEAN/STD DEV per device.

For the average learning curves acquired for different modalities, there is significant effect, F(2,6.38) = 4.36, P = 0.0286. Paired analyses reveal that the mean learning time per hit for the static posture is significantly less than that of the haptic 3D mouse modality, F(1,12.72) = 5.89, P = 0.0381.

B. *Extra Observations*

Figures 15 and 16 present some more feedbacks from participants regarding their preferences in combination of devices to be used for this game application (the right balance of input/output methods), and their computer skills.

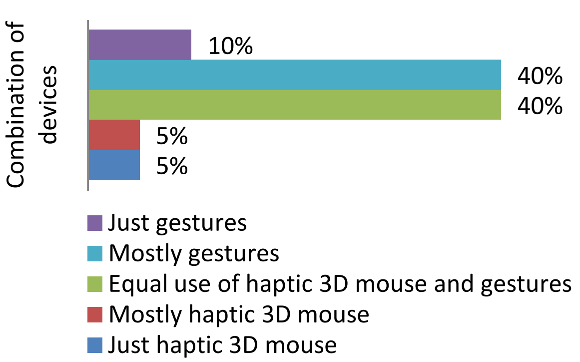


Figure 15. Users' preferences in combining the devices for this application.

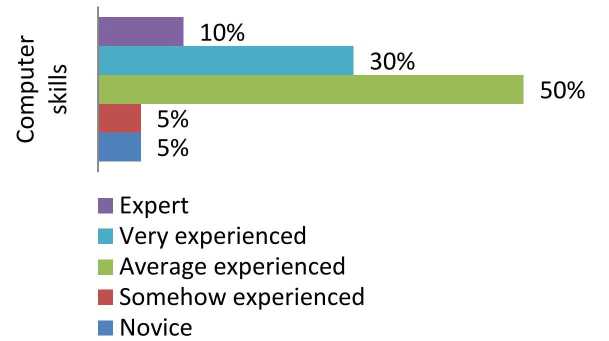


Figure 16. Participants' computer skills.

C. *Discussions*

Computer skills of participants widely ranged from novice to expert. As a result, it would be valid to conclude that the findings of this study are extensively applicable for practical purposes and not just tech-savvy users. Furthermore, the wide age span of the participants ensures that the findings are comprehensively general and age-independent.

According to the provided statistical analyses, we can summarize that static gestures are shown to be faster and easier than dynamic gestures while being more precise, less fatiguing, more natural, and more pleasant than both other modalities. It is faster and lighter to perform hand and arm static postures due to fewer movements involved in the process compared to dynamic gestures. As a result, they would be less fatiguing, more pleasant, and more precise.

Furthermore, continuous attachment to the haptic 3D mouse reduces the naturalness of experiences through the 3D virtual environment, while the ergonomic design and force feedback increase fatigue and error rate. The haptic 3D mouse is easier than dynamic gestures meanwhile having less mobility (space of interaction) than both other modalities. While the easiness factor was discussed earlier, the mobility can be argued in terms of the stationary nature and spatial boundaries of the haptic 3D mouse.

Overall, static postures are directly and indirectly more satisfactory than the other two modalities. Since most of the human factors showed to be superior for static postures compared to the other modalities, it is valid to expect higher rates for the direct overall satisfaction feedback as well. The weighted overall satisfaction is in complete correlation with the direct method. This indicates that intuitive feedback regarding satisfaction on a modality is a reliable means for design evaluation.

The experiments also reveal various conclusions regarding primitive tasks, namely moving the pointer, grabbing the ball, pulling the ball, and finally throwing the ball. It is shown that static postures application is more approved for moving the pointer compared to dynamic gestures application while maintaining superiority over the other two modalities when grabbing and pulling the ball. Finally, it is more difficult to grab and throw the ball with dynamic gestures compared to the other two modalities.

None of the users were familiar in applying any of these three modalities before this experiment. As a result, the obtained learning curves can be valid measures for further investigation of the modalities. The acquired learning curves indicate that the mean learning time per hit for the static posture is less than that of the haptic 3D mouse modality. This can be interpreted based on at least the two factors of naturalness and fatigue which influence the learning capabilities of participants.

According to Figure 15, direct feedback from participants demonstrates that most of them suggested "equal use of haptic 3D mouse and gestures" with "mostly gestures" as their preferred combination of modalities.

V. Conclusion

A 3D slingshot game was implemented using XNA, OpenNI, NITE, and OpenCV. Three modalities were defined using a haptic 3D mouse and Kinect. Two types of vision-based input methods were developed for Kinect as static and dynamic gestures. User experiments were conducted to study the different human factors associated with modalities. Precision (error) and efficiency (time) along with satisfaction criteria such as ease-of-use, fun-to-use, fatigue, naturalness, and mobility were rated in each modality and ranked independently. Static postures proved to be most efficient, precise, fun, and natural to use compared to the other modes. Furthermore, overall satisfaction was also acquired as a direct feedback from participants. Alternatively overall satisfaction was also computed by integrating the satisfaction criteria's ranks in their rates. The result of the computed overall satisfaction showed to be in complete conformity with the direct satisfaction ratings, yielding that intuitive feedback on satisfaction can be valuable means for design studies. Overall, static postures are directly and indirectly more satisfactory than the other two modalities. In terms of learning to utilize the modalities, static postures once more showed superiority against the others. Finally, it should be mentioned that even though static postures maintain superiority in many aspects over dynamic gestures and haptic 3D mouse, the latter two modalities cannot be completely ignored from being incorporated in HCI systems. This is because of the fact that dynamic gestures and haptic 3D mouse provide a vaster domain for possible gesture selection and real 4D tasks (e.g., precise following of a trajectory).

References

[1] F. Farhadi-Niaki, R. Ghasem-Aghaei, and A. Arya, "Empirical study of a vision-based depth-sensitive human-computer interaction system," *10th Asia Pacific Conference on Computer Human Interaction*, Matsue, Japan, pp. 101-108, 2012.

[2] A. Jaimes and N. Sebe, "Multimodal human computer interaction: a survey," *Computer Vision and Image Understanding*, 108(1-2), pp. 116-134, 2007.

[3] G. Ye, "Applying vision to intelligent human-computer interaction," Ph.D dissertation, Department of Computer Science, The Johns Hopkins University, 2005.

[4] J. Nespoulous, P. Perron, and A. R. Lecours, *The biological foundations of gestures: motor and semiotic aspects*, Hillsdale, MJ: Lawrence Erlbaum Associates, 1986.

[5] A. Kendon, *Current issues in the study of gesture. In the biological foundation of gestures: motor and semiotic aspects*, Hillsdale: Lawrence Erlbaum Assoc., pp. 23-47, 1986.

[6] V. Pavlovic and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," *The IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):667, 1997.

[7] R. Vatavu, S. Pentiuc, and C. Chaillou, "On natural gestures for interacting in virtual environments," *Advances in electrical and computer engineering*, 24(5), 2005.

[8] B. Yoo, J. Han, C. Choi, H. Ryu, D. S. Park, and C. Y. Kim, "3D remote interface for smart displays," *The ACM conference on Human Factors in Computing Systems*, Vancouver, Canada, 2011.

[9] N. Villaroman, D. Rowe, and B. Swan, "Teaching natural user interaction using OpenNI and the Microsoft Kinect sensor," *The ACM Special Interest Group on Information Technology Education*, West Point, New York, USA, 2011.

[10] M. Bhuiyan and R. Picking, "A gesture controlled user interface for inclusive design and evaluative study of its usability," *Journal of Software Engineering and Applications*, 2011.

[11] L. C. Ebert, G. Hatch, G. Ampanozi, M. J. Thali, and S. Ross, "You can't touch this: touch-free navigation through radiological images," *Special Area Group of Educators*, 2011.

[12] M. A. Srinivasan, *What is haptics?*, [Online]. Available: http://www.sensable.com/documents/documents/what_is_haptics.pdf [Accessed: April 2012].

[13] Sensible Technologies, *The Phantom devices*, [Online]. Available: http://www.sensable.com/products/phantom_ghost/phantom-omni.asp [Accessed: April 2012].

[14] Immersion Corporation, *The Cyber Grasp*, Cyber Touch, Cyber Glove devices, [Online]. Available: http://www.immersion.com/3d/products/cyber_grasp.php [Accessed: April 2012].

[15] V. Hayward and M. Cruz-Hernandez, "Tactile display device using distributed lateral skin stretch," *In Proc. 8th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, vol. 2, pp. 1309–1314, 2000.

[16] C. Wagner, S. Lederman, and R. Howe, "A tactile shape display using RC servomotors," *In Proc. 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator*, pp. 354–355, 2002.

[17] M. Cavusoglu, D. Feygin, and F. Tendick, "A critical study of the mechanical and electrical properties of the Phantom haptic interface and improvements for high performance control," *Presence: Teleoperators and Virtual Environments*, 11(6):555–568, 2002.

[18] T. Massie and J. Salisbury, "The Phantom haptic interface: a device for probing virtual objects," *In Proc. Symposium on Haptic Interfaces for Virtual Environment and Teleoperator*, 1994.

[19] OpenNI Organization, *Documentation: Programmer's Guide*, OpenNI™, 2010. [Online]. Available: http://www.openni.org/Documentation/ProgrammerGuide.html [Accessed:Feb.2012].

[20] PrimeSense Inc., *Prime Sensor™ NITE 1.3 Controls Programmer's Guide*, PrimeSense Inc., 2010. [Online]. Available: http://www.andrebaltazar.files.wordpress.com/2011/02/nite-controls-1-3-programmers-guide.pdf [Accessed: Feb. 2012].

[21] S. W. Huck, *Reading statistics and research*, 3rd ed., New York: HarperCollins, 2000.

Interactive Dynamic Simulations with Co-Located Maglev Haptic and 3D Graphic Display

Peter Berkelman
Department of Mechanical Engineering
University of Hawaii at Manoa
Honolulu Hawaii, USA
Email: peterb@hawaii.edu

Sebastian Bozlee
Mathematics Department
Unversity of Portland
Portland Oregon, USA
Email: sebbyj@gmail.com

Muneaki Miyasaka
Department of Mechanical Engineering
University of Washington
Seattle Washington, USA
Email: muneaki@uw.edu

***Abstract*—We have developed an system which combines realtime dynamic simulations, 3D display, and magnetic levitation to provide high-fidelity co-located haptic and graphic interaction. Haptic interaction is generated by a horizontal array of cylindrical coils which act in combination to produce arbitrary forces and torques in any direction on magnets fixed to an instrument handle held by the user, according to the position and orientation sensed by a motion tracking sensor and the dynamics of a realtime physical simulation. Co-located graphics are provided by a thin flat screen placed directly above the coil array so that the 3D display of virtual objects shares the same volume as the motion range of the handheld instrument. Shuttered glasses and a head tracking system are used to preserve the alignment of the displayed environment and the interaction handle according to the user's head position. Interactive demonstration environments include rigid bodies with solid contacts, suspended mass-spring-damper assemblies, and deformable surfaces.**

***Keywords*-haptics, interaction, simulation**

I. Introduction

The ideal of virtual reality and haptic interfaces is to physically interact with simulated objects with the highest possible fidelity in both the graphical display and the kinesthetic forces and torques sensed by the user during interaction. Computer-generated graphics can produce highly realistic, dynamic 3D imagery in real time, but haptic interfaces are generally based on single point contact feedback, tactile cues, and linkage devices which have various limitations in their force and motion ranges, frequency response bandwidth, and resolution.

Our system combines a graphical display with a large range of motion magnetic levitation device, as shown in Fig. 1. The graphical display is placed directly above a horizontal array of cylindrical coils and underneath the instrument handle held by the user, so that electromagnetic forces and torques can be generated on magnets embedded in the handle as the instrument is moved by the user into contact with the displayed simulated environment.

The magnets in the instrument handle, the coil array with its current amplifiers, and the motion tracking sensor with its infrared LED markers, function together as a magnetic levitation system. Its motion range is approximately 100x100 mm horizontally, up to 75 mm vertically, with unlimited yaw and tilt up to 40 degrees.

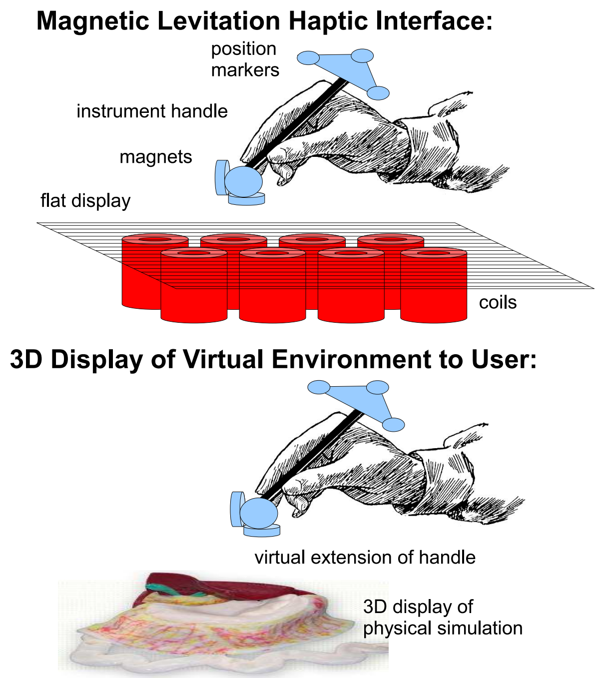


Figure 1. Co-Located Maglev Haptic and 3D Graphic Display

A secondary, slower and less precise motion tracking system tracks the position of the user's head so that the 3D views are generated correctly according to the position of each of the user's eyes. A pair of shuttered glasses, synchronized to the update rate of the graphics on the monitor, is worn by the user so that each eye sees a different image as the shutters alternate. In practice, this head tracking system allows the user to observe the handheld instrument and the 3D displayed environment together from the side and from above, in a natural ergonomic position for hand-eye coordination during dextrous manipulation of a handheld instrument or tool. Examples of relevant dextrous

 ISBN: 978-1-61208-250-9

tool manipulation tasks include any writing, carving, or cutting tasks, operation of wrenches or screwdrivers, and medical needle manipulation for suturing, injections, and biopsy.

The real-time haptic interaction and graphical display are generated from a dynamic simulation which must perform collision detection, finite element deformation, and haptic rendering sufficiently quickly to support graphical updates at 30-60 Hz and haptic interaction and magnetic levitation at 800-1000 Hz. These tasks are sufficiently computationally intensive to be the limiting factor regarding the resolution and realism of the simulated environment.

II. Background

The realization of our interactive system depends on the performance and integration of technology in the areas of maglev haptics, graphics, and physical simulation. Relevant prior work in each of these areas is surveyed below.

A. Co-Located Haptics and Graphics

3D graphics and haptic force and/or torque feedback can be generated at the same location by simply placing the 3D display behind the haptic interaction device, however this method has two drawbacks. First, the body of the haptic interface device partially occludes the display, and second, there may be a significant difference produced between the perceived location of the displayed imagery and the surface of the screen, so that the convergence and focal distance of the user's eyes do not match, which is unnatural and may cause discomfort to the user.

ReachIn, *Immersivetouch*, and *SenseGraphics* systems [1] use a partially silvered mirror between the head and hand of the user, so that the display can be moved out of the way and the focal and convergence distances of the user's eyes can be matched. The haptic device and the user's hand do not occlude the 3D graphics behind them, but rather the real and virtual environments are superimposed and semitransparent due to the half-silvered mirror, which may be a distraction to the user.

The "what you see is what you feel" system [2] uses a thin flat display with a camera behind it. The video image of the user's hand is then extracted from the camera view using a green screen chroma-key technique, and rendered in the virtual environment. Holographic display [3] is another method which has been used for haptic and graphic co-location.

Comparative studies have shown [4], [5] evidence of improved perception and performance from co-located haptic interaction.

B. Haptic Magnetic Levitation

Hollis and Salcudean first applied Lorentz force magnetic levitation devices [6] to haptic interaction and force-feedback teleoperation. Lorentz force magnetic levitation haptic interaction development continued with other more specialized device designs [7] [8] and larger range devices developed by Berkelman [9] [10].

The design and function of the magnetic levitation system used here is described in [11]. This system uses a fixed planar array of cylindrical coils to levitate a platform of one or more cylindrical magnets. The yaw of the levitated platform is unlimited and its horizontal motion range is determined by the size of the planar array. Vertical levitation distances of up to 75 mm and tilt angles of 45 degrees are achievable, depending on the mass of the levitated platform and the dimensions of the magnets used.

Similar tabletop-scale large range magnetic levitation systems have been developed for suspension of models in wind tunnels [12] and for micromanipulation using pole pieces to shape magnetic fields [13].

C. Realtime Physical Simulation

Realistic software simulations of dynamic physical environments have been developed by Baraff both for rigid [14] and deformable [15] objects, including efficient collision and reaction force detection and surface friction. Freely available physical simulation software packages include the SOFA framework [16] [17], Bullet Physics, and the PhysX library from NVIDIA. Higher resolution and performance can be obtained by using precomputed deformation modes [18] and 6-DOF haptic rendering including torque feedback as well as force on an interactive instrument can be integrated with simulations [19].

Several software packages are freely available for haptic rendering and realtime physical simulation. H3D [20] and Chai3D [21] include driver interfaces for common commercial haptic interface devices such as the Sensable Technologies Phantom [22]. A programming interface is also available with the magnetic levitation haptic interface from Butterfly Haptics LLC [23].

III. Implemented System

The motion tracking, magnetic levitation control, haptic rendering, physical simulation, and graphical display in our current system are all executed in real time in separate threads on a single quad-core PC in Linux 2.6. GNU C/C++ was used for all programming. An initial demonstration concept of the system with a simulation of a single paddle instrument and a ball rolling on a plane, an earlier magnet and coil configuration, and a conventional 2D display was demonstrated previously [24]. The current system is shown in Fig. 2, including the planar 3D display, haptic instrument handle, current amplifiers, and head tracker.

A. Magnetic Levitation

The motion tracking of the handheld instrument in our system is done using a Northern Digital Optotrak Certus position sensor and three infrared Smart Markers. Motion

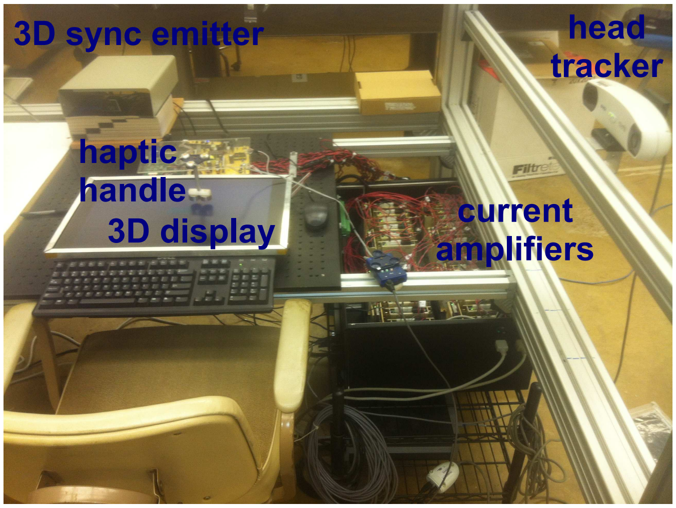


Figure 2. Implemented system

tracking updates are provided at 860 Hz with a position resolution of approximately 0.01 mm for each marker. Actuation forces and torques are generated by a closely packed array of 27 cylindrical coils, each with 1000 windings, 25 mm diameter, and 30 mm height. Either a two-magnet or four-magnet instrument handle can be used with the system; the two-magnet 125 g instrument can provide greater haptic forces and torques but is more massive and bulky, and the smaller 75 g four-magnet instrument occludes the user's view of the display less due to its compact size. Forces are limited to approximately 4 N due to heating of the actuation coils, although higher momentary peak forces are possible.

The four-magnet instrument is shown in Fig. 3 levitated above the 27-coil array at a height of 30 mm and a tilt angle of 20 degrees. This coil array is underneath the 3D display monitor shown in Fig. 2. The motion tracker for the haptic instrument is mounted on a rigid frame at ceiling level, looking downwards.

The design and evaluation methods used in the development of the magnetic levitation system are described in [25].

B. 3D Display

The NVIDIA 3D Vision package was used with Linux drivers to provide 3D display of the simulated environment. This package uses shutter glasses which are synchronized with the graphics card by an infrared emitter box. A Quadro 4000 graphics card was used with a ViewSonic vm2268 monitor with a 120 Hz update rate. OpenGL and GLUT graphics libraries are used for the 3D graphics rendering.

The case of the monitor was removed and backplane circuit boards and wiring were moved so that the monitor backlight and display could be placed directly on the coil array. The combined thickness is under 10 mm, so that haptic forces and torques can be applied to the handheld instrument up to a vertical height of at least 60 mm. Magnetic fields

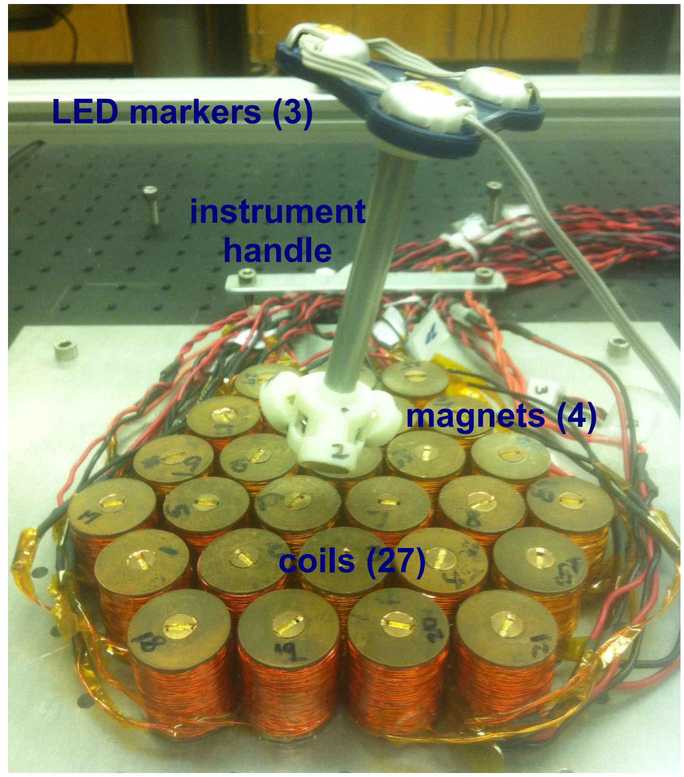


Figure 3. Levitated 4-magnet instrument

from the instrument magnets and coil array were not found to interfere with the display, and there are no ferromagnetic components in the display to interfere with the magnetic levitation system. A thin sheet of plastic was placed on the monitor screen for protection from impacts from the magnets and instrument.

Head tracking was implemented using a Northern Digital Polaris Vicra and passive reflective markers to produce correct 3D display according to the position of each eye. The spatial position and orientation of the shutter glasses from the positions of four reflective markers fixed to the glasses. Position and orientation data were updated at a 10 Hz rate with a resolution of approximately 0.1 mm for each marker. It would be possible to track both the magnet instrument and and the user's head using a single motion tracking system, but this would require using wired infrared markers on the 3D shutter glasses, slowing down the update rate of the magnetic levitation localization due to the additional LED markers on the glasses, and mounting the localizer at least 3.5 m high so that its sensing volume includes the location of the glasses.

As both the Optotrak and Polaris motion trackers use infrared position sensing, and 3D Vision systems uses infrared communication to synchronize display frames with the shutter glasses, it is necessary to ensure that each infrared system does not interfere with the others. In our

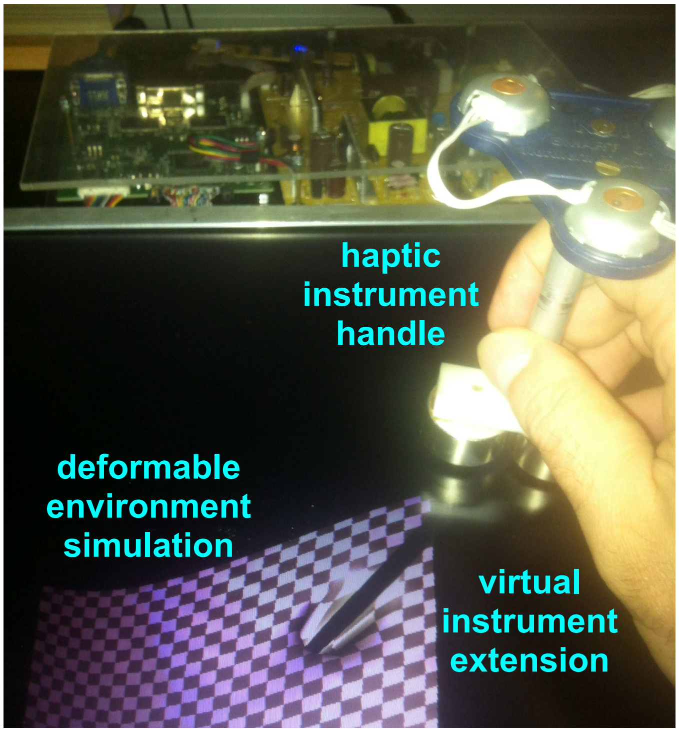


Figure 4. Deformable simulation

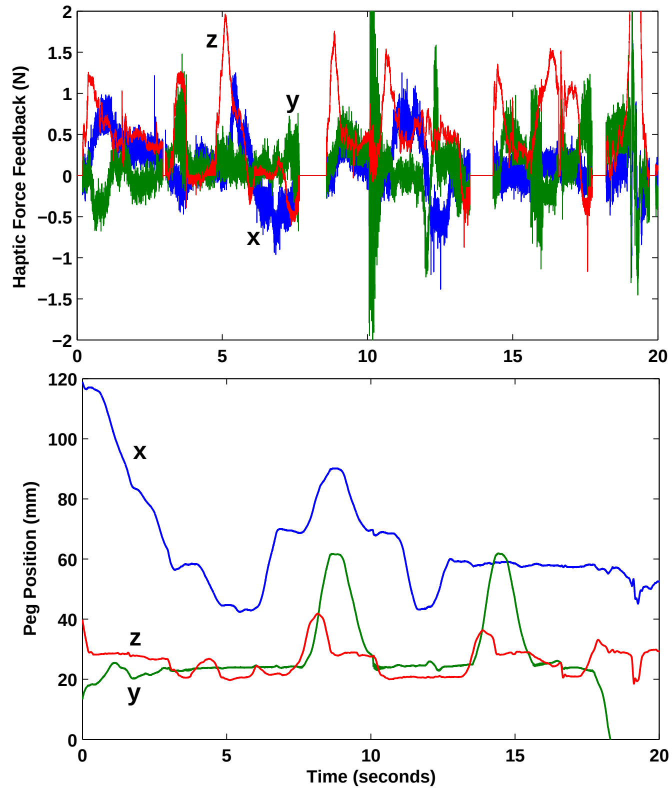


Figure 5. Interactive peg-in-hole simulation

system, each set of emitters and receivers are oriented in orthogonal directions and positioned so that each emitter is only visible to its corresponding receiver. The Optotrak sensor is mounted above the table looking down at the LEDs on the instrument, the Polaris is mounted on the side of the table to track the reflective markers on the side of the glasses, and the synchronization emitter is mounted at the front of the tabletop.

C. Simulations

Basic interactive simulations which have been implemented on our system at present include point, edge, and face contacts between simple solid shapes such as square peg-in-hole insertion, simple dynamic environments including suspended masses and springs, and rolling objects. These simple initial simulations allow the dynamics and contact models of the environments to be modified and adjusted to provide the most realistic haptic interaction while preventing unstable dynamics.

A more sophisticated simulation which involves an instrument contacting a deformable surface is shown in Fig. 4. In this simulation, a virtual extension is added to the actual haptic instrument handle, and the deformation of the surface and reaction forces and torques on the instrument are calculated at the haptic update rate. Damping is added to the internal dynamics of the deformable body and the surface dynamics during contact with the haptic instrument.

The MLHI library and programming interface, originally from Butterfly Haptics, has been adapted for use with our system, and can be used for haptic rendering and communication between simulation and magnetic levitation threads with a haptic update rate of 1000 Hz. Alternatively, haptic rendering and dynamic simulation calculations can be performed synchronously with the motion tracking at 860 Hz.

IV. Results

Force and position experimental data in x, y, and z directions obtained during interactive simulations are presented in Figs. 5 and 6. The position data was measured by the position tracking system, and the force data are calculated by the simulations and generated by the coil array of the magnetic levitation system in real time. The commanded forces were shown to be within 0.1 Newtons of force sensor measurements throughout the range of the magnetic levitation system in [11].

The Fig. 5 plots are from a haptic peg-in-hole simulation in which a 25 mm square peg is controlled by the haptic instrument handle and inserted into a 27x54 mm, 10 mm deep square hole. The Fig. 6 plots are from a deformable simulation in which a pointed virtual instrument contacts a deformable object, as shown in Fig. 4. For both cases, haptic forces and torques are zero while the instrument is moving freely, contact forces are approximately proportional to the

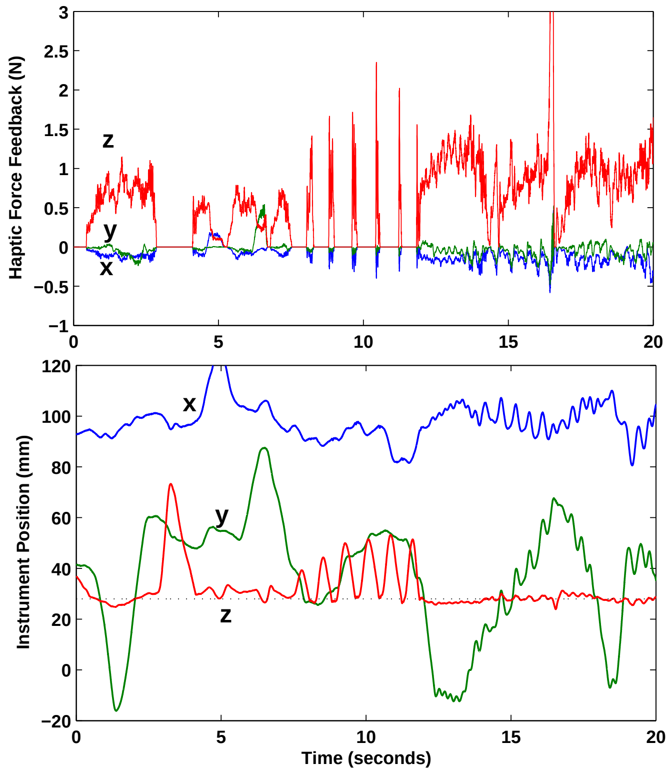


Figure 6. Interactive palpation of deformable surface

depth of contact, and haptic torques depend on each contact force and the displacement between the contact point and the center of the haptic instrument and simulated tool.

In the peg-in-hole simulation of Fig. 5, the peg is not in contact with the hole or top surfaces at the 8-9 and 14-15 second intervals, the z coordinate is greater than 30, and there is no haptic force feedback. As the peg is moved in and out of the hole, the z position moves between 20 and 30 mm. The x position can vary between approximately 40 and 70 mm while the peg is in the hole, as the hole is more than twice as wide as the peg in the x direction. Non-zero x and y forces are present when the virtual peg is pushed against any of the four sides of the virtual hole. Contact stiffnesses are approximately 0.4 N/mm and the kinetic and static friction coefficents are 0.15 in the simulations.

For the deformable surface of Fig. 6, the probe is moved across the surface during the 12-20 second interval, and the surface is struck with the probe several times from 8-12 seconds. The object was modeled with millimeter-scale surface variations rather than a smooth flat surface. This surface texture therefore produces variable vertical (z) forces in response to horizontal (x and y) motions of the intstrument tip. Oscillations in both the position and force data can be seen in the 12-20 second period due to sticking and slipping of the sliding surface contact. Overall the force plots are smoother in the deformable surface simulation than the peg-in-hole simulation due to the compliance and lower friction of the deformable surface.

V. Further Work

At present, the magnetic levitation and motion tracking aspects of our system are fully developed, but the interactive environments are at a preliminary stage. We plan to refine the detail and physical realism of the simulated environments to a degree where they are useable and provide measureable benefits in medical training tasks such as surgery, intubation, and needle driving. User studies will be conducted to evaluate the benefits of colocated haptic and graphical training of simulated medical procedures.

The complexity of the modelled environment and the sophistication of the simulated dynamics can be improved by using the graphics processor for additional numerical computations, as a general purpose graphics processing unit or GPGPU. NVIDIA provides the CUDA [26] programming interface to utilize the parallel processing capabilities of the GPGPU on the graphics cards used.

One more planned improvement to be made on the system is to reconfigure the system to be simpler and more compact.

VI. Conclusion

Our co-located haptic and graphic interface system is novel in that there is no hardware between the user and the display other than the handheld interaction instrument. The 3D environment is displayed close to the surface of the monitor, so there is no conflict between visual convergence and focal ranges. Electromagnetic force and torque actuation is used for haptic interaction rather than a motorized linkage, providing advantages in backdriveability, precision, and response frequency bandwidths.

We have demonstrated the feasibility and function of our system with the basic simulation environments described.

Acknowledgment

This work was supported in part by National Science Foundation grants IIS-0846172 and CNS-0551515, and by the University of Hawaii College of Engineering.

References

[1] C. Luciano, P. Bannerjee, L. Florea, and G. Dawe, “Design of the ImmersiveTouch: a high-performance haptic augmented virtual reality system,” in *11th International Conference on Human-Computer Interaction*, Las Vegas, August 2005.

[2] Y. Yokokoji, R. Hollis, and T. Kanade, “WYSIWYF display: A visual/haptic interface to virtual environment,” *Presence*, vol. 8, no. 4, pp. 412–434, August 1999.

[3] P. Olsson, F. Nysjo, S. Siepel, and I. Carlbom, “Physically co-located haptic interaction with 3d displays,” in *IEEE Haptics Symposium*, Vancouver, March 2012, pp. 267–272.

[4] D. Swapp, V. Pawar, and C. Loscos, “Interaction with haptic feedback and co-location in virtual reality,” *Presence*, vol. 10, no. 1, pp. 24–30, April 2006.

[5] G. Jansson and M. Ostrom, "The effects of co-location of visual and haptic space on judgements of form," in *Euro-Haptics*, Munich, June 2004, pp. 516–519.

[6] R. L. Hollis and S. E. Salcudean, "Lorentz levitation technology: a new approach to fine motion robotics, teleoperation, haptic interfaces, and vibration isolation," in *Proc. 6th Int'l Symposium on Robotics Research*, Hidden Valley, PA, October 1993.

[7] S. Salcudean and T. Vlaar, "On the emulation of stiff walls and static friction with a magnetically levitated input-output device," in *ASME IMECE*, Chicago, November 1994, pp. 303–309.

[8] P. J. Berkelman, R. L. Hollis, and S. E. Salcudean, "Interacting with virtual environments using a magnetic levitation haptic interface," in *Int'l Conf. on Intelligent Robots and Systems*, Pittsburgh, August 1995.

[9] P. J. Berkelman and R. L. Hollis, "Lorentz magnetic levitation for haptic interaction: Device design, function, and integration with simulated environments," *International Journal of Robotics Research*, vol. 9, no. 7, pp. 644–667, 2000.

[10] P. Berkelman and M. Dzadovsky, "Extending the motion ranges of magnetic levitation for haptic interaction," in *Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, Salt Lake City, March 2009, pp. 517–522.

[11] ——, "Magnetic levitation over large translation and rotation ranges in all directions," *IEEE/ASME Transactions on Mechatronics*, vol. 18, no. 1, pp. 44–52, 2013.

[12] N. J. Groom and C. P. Britcher, "A description of a laboratory model magnetic suspension testfixture with large angular capability," in *IEEE Conference on Control Applications*, Dayton, September 1992, pp. 454–459.

[13] M. B. Khamesee and E. Shameli, "Regulation technique for a large gap magnetic field for 3d non-contact manipulation," *Mechatronics*, vol. 15, pp. 1073–1087, 2005.

[14] D. Baraff, "Interactive simulation of solid rigid bodies," *IEEE Computer Graphics and Applications*, vol. 15, pp. 63–75, 1995.

[15] D. Baraff and A. Witkin, "Dynamic simulation of non-penetrating flexible bodies," in *Computer Graphics* (*Proc. SIGGRAPH*), vol. 26. ACM, July 1992, pp. 303–308.

[16] J. Allard, S. Cotin, F. Faure, P.-J. Bensoussan, F. Poyer, C. Duriez, H. Delingette, and L. Grisoni, "Sofa an open source framework for medical simulation," in *Medicine Meets Virtual Reality (MMVR'15)*, Long Beach, USA, February 2007.

[17] M. Marchal, J. Allard, C. Duriez, and S. Cotin, "Towards a framework for assessing deformable models in medical simulation," in *International Symposium on Computational Models for Biomedical Simulation*, London, July 2008.

[18] J. Barbic and D. James, "Six-dof haptic rendering of contact between geometrically complex reduced deformable models," *IEEE Transactions on Haptics, preprint published online*, June 2008.

[19] D. James and D. Pai, "A unified treatment of elastostatic and rigid contact simulation for real time haptics," *Haptics-e*, vol. 2, no. 1, 2001.

[20] *HAPI Manual*, 1st ed., SenseGraphics Inc., May 2009.

[21] Http://www.chai3d.org/.

[22] T. Massie and K. Salisbury, "The PHANToM haptic interface: A device for probing virtual objects," in *Proceedings of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, Chicago, Illinois, November 1994.

[23] *Magnetic Levitation Haptic Interface API Reference Manual*, 1st ed., Microdynamic Systems Lab, Carnegie Mellon University, September 2008.

[24] P. Berkelman, M. Miyasaka, and J. Anderson, "Co-located 3d graphic and haptic display using electromagnetic levitation," in *IEEE Haptics Symposium*, Vancouver, March 2012, pp. 77–81.

[25] P. Berkelman and M. Dzadovsky, "Novel design, characterization, and control method for large motion range magnetic levitation," *IEEE Magnetics Letters*, vol. 1, January 2010.

[26] *Nvidia CUDA Compute Unified Device Architecture Programming Guide 1.0*, Nvidia, 2007.

 ISBN: 978-1-61208-250-9

Haptic System for Eyes Free and Hands Free Pedestrian Navigation

Nehla Ghouaiel
ESTIA-IRIT
F-64210 Bidart, France
n.ghouaiel@estia.fr

Jean-Marc Cieutat
ESTIA-IRIT
F-64210 Bidart, France
j.cieutat@estia.fr

Jean-Pierre jessel
IRIT
UMR 5505, F-31062 Toulouse, France
jean-pierre@irit.fr

Abstract—**Until now, Augmented Reality was mainly associated with visual augmentation which was often reduced to superimposing a virtual object on to a real world. We present in this document a vibro-tactile system called HaptiNav, which illustrates the concept of Haptic Augmented Reality. We use the haptic feedback method to send users information about their direction, thus enabling them to reach their destination. To do so, we use a turn by turn metaphor which consists of dividing the route into many reference points. In order to assess the performances of the HaptiNav system, we carry out an experimental study in which we compare it to both Google Maps Audio and Pocket Navigator systems. The results show that there is no significant difference between HaptiNav and Google Maps Audio in terms of performance, physical load and time. However, statistical analysis of the mental load, frustration and effort highlights the advantages of HaptiNav compared to two other systems. In the light of the results obtained, we present possible improvements for HaptiNav and describe its second prototype, at the end of this paper.**

Keywords; haptic navigation; augmented reality; mobile computing; human computer interaction.

I. INTRODUCTION

Until now Augmented Reality was mainly associated with visual augmentation, which is often reduced to superimposing a virtual object onto a real object [5]. However, the concept of Augmented Reality is not limited to sight and could be extended to other senses, ie., hearing, smell and touch. Panagiotis et al. [10] classed user experience input and output, in an augmented reality environment. The input may be audio, visual, tactile or kinaesthetic, whilst the output may only be visual, haptic or audio. In B.Bayart [6] presented three different existing taxonomies for Augmented Reality, and studied their direct extension in terms of augmented haptics. As a continuation, the Haptic Augmented Reality taxonomy was introduced and separated into two categories: augmented haptics and haptic augmentation. As with the classification of Fuchs et al. [4], Haptic Augmented Reality systems can be used either to augment existing data or to add information, referred to as enhanced haptics and haptic enhancing respectively. Enhanced haptics is defined as when the haptic modality amplifies or modulates a haptic datum sent back to the user. In some applications, it may be important to be able to touch data which are not on a human scale and which are not perceptible by direct contact with one of the body parts. Thus, feeling holes and bumps which are no larger than on a mesoscopic scale, a sort of haptic microscope, is an example of this. The concept of haptic enhancing can be summarized as scenarios where the haptic modality is used to send additional information to the user. Some researchers have explored the possibilities of transferring emotions through haptic interfaces. Shneiderman [13] defined a (computer) icon as an image, a drawing or a symbol representing a concept, and in 2004 S. A. Brewster et al. [7] introduced the notion of tacton or tactile icon which is like visual icons, represents a tactile concept. After analyzing the possibilities provided by tacton vibrations in [8], the work presented in [16], endeavors to go further by trying to simulate emotions through tactile vibrations.

In the study presented in this paper, we explore how to use Augmented Reality in its haptic modality, in order to guide pedestrian in a new urban environment. In this context, several questions are raised: Is the haptic modality efficient enough to guide pedestrians? Is it robust enough to allow hands free and eyes free navigation? To answer these questions we implement a vibro-tactile system which illustrates the concept of haptic augmented reality. We use the haptic modality to send users informations about their directions, enabling them to reach their destinations. In the first Section of this paper, we describe the prototype developed for HaptiNav. We then present the software structure of HaptiNav and the algorithm used. In Section 4, we detail the experimental study carried out to assess performances of HaptiNav system, in comparison with the standard Google Maps Audio and with another vibro-tactile system. All of these systems were tested without any visual support. At the end of this paper, we present possible improvements for HaptiNav and describe the second prototype for our system.

II. DESIGN OF THE HAPTIC INTERFACE

The Figure 2 highlights three possible prototypes for our system. The selection criterion consists of finding the prototype which provides user friendly navigation, which enable users to navigate hands free and eyes free.

We carry out a comparative study to choose the prototype which best corresponded to our selection criterion. The first line in Figure 2, shows a tactile tablet designed to be placed in the palm of the hand. Jin et al. [9] created a tactile tablet made up of 12 panels, each panel contains a tactor (ie., a vibrating motor). Their tactile tablet contains 12 vibrators, forming a 4 lines and 3 columns matrix. In order to send to user spatial and directional information, T-mobile system [9] combines three vibrators. For instance, T-mobile system makes tactors vibrate in the first line

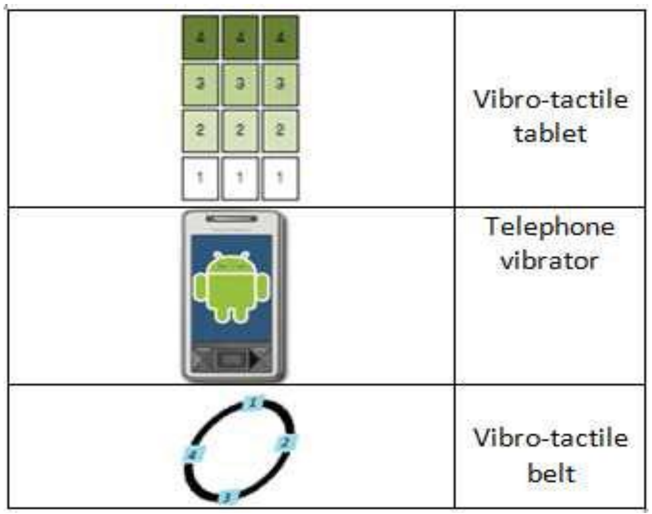


Figure 1. Possible prototypes.

to show the north. The main disadvantage of this prototype is that it needs to be held in one of the user's hands, therefore it prevents hands free navigation. The second prototype is based on a single vibrator integrated in a mobile phone. The Pocket Navigator system [3] uses this prototype; it codes the direction which must be followed by the user in different vibration modes, known as tactons. The approach presented in Pocket Navigator uses three different rhythms to tell the user to go straight on, turn right, turn left or stop. As shown in the figure below, the system translates the action of going straight on by two short successive pulses, turning left by a long vibration followed by a short one, turning right by a short vibration followed by a long one and turning back by three short pulses.

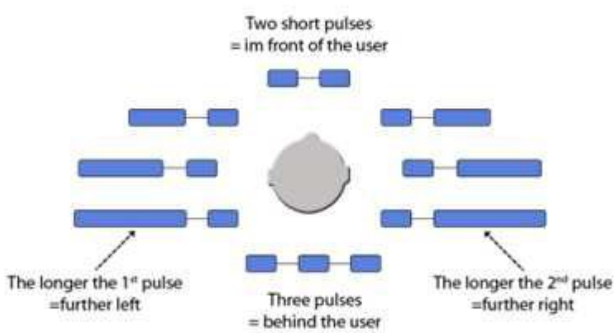


Figure 2. Pocket Navigator tactile diagram [3].

We chose to develop the vibro-tactile belt prototype, namely to make navigation more user-friendly. As previously highlighted, we want users to be able to reach their destinations without using vision or hands. The aim behind this choice is to enable users to concentrate on road traffic and obstacles, rather than on the navigation system. For people visiting a town for the first time, the advantage is that it guides them whilst at the same time, allowing them to fully concentrate on the new environment. Unlike ActiveBelt [15], our system has four vibrators.

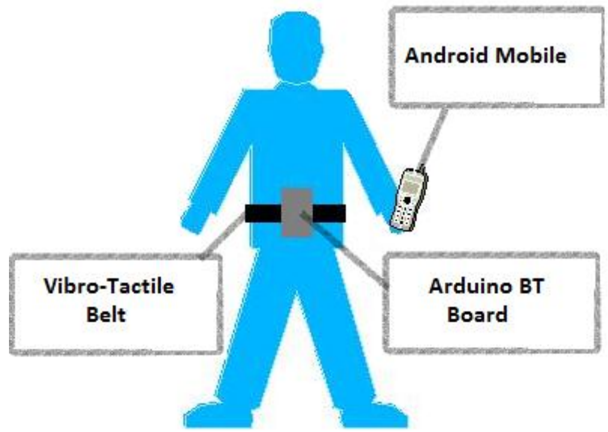


Figure 3. Prototype of HaptiNav.

III. IMPLEMENTATION

Our system called HaptiNav, consists of an Android Smartphone Galaxy S2 and an Arduino BT electronic board [1] as shown in Figure 4. Arduino BT is an Arduino board with an integrated Bluetooth module, thus enabling wireless communication with the smartphone. In order to implement the system's applicative aspect (deployed on the mobile and the Arduino board), we use Android SDK programming interface (API) and Arduino software.

The Android development environment consists of the Android development tool (ADT) integrated in Eclipse. We use Arduino freeware to develop the application loaded on the Arduino BT board, which controls the vibrators and the microcontroller. We chose this software because it is the proprietary platform of the Arduino electronic card.

To make the mobile application and the Arduino sketch communicate, we use Amarino software interface. It is developed as part of the Android meet Arduino project [2]. Amarino was launched by Bonifaz Kaufmann [2] in 2009, and developed at the University of Klagenfurt in Austria. There is another tool which enables Android and Arduino to communicate via USB: ADK (Android Open Accessory Development Kit). We are unable to adopt this solution since it is only available for Android version 13. However, Galaxy S2 has Android version 11. The diagram below shows the structure of the applicative part of our system. It is divided into three modules: Android, Arduino and Android Arduino interface. The Android application constantly calculates the difference between the user's orientation and the orientation of the route's closest way point. The Android application sends difference in orientation angle to the Arduino sketch through the Amarino plug-in. Consequently, the Arduino sketch activates the belts vibrator corresponding to the direction sent.

IV. ALGORITHM IMPLEMENTED

Unlike the compass metaphor used in ActiveBelt [11], we use another metaphor which we shall refer to it as turn by turn metaphor. With the compass metaphor, the difference between the user's current orientation and the destination's orientation is

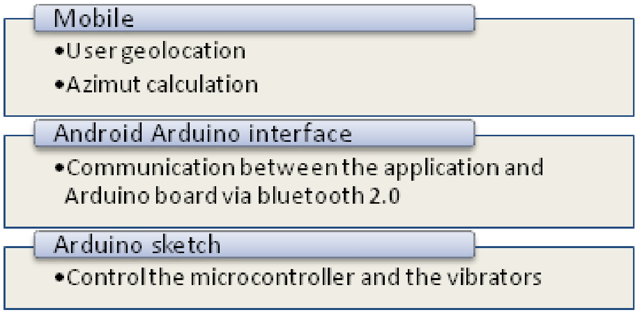


Figure 4. Structure of the system's applicative part.

instantly calculated and then, the user is constantly redirected towards the final destination. The turn by turn metaphor consists of dividing the route into many reference points. Thus, the system shows the user which direction to take for each way point. We refer to the user's current orientation as OC and the desired orientation as OD. The desired orientation OD is extracted from the KML file (Keyhole Markup Language), generated by Google Maps. Desired orientation can be found in the tag <heading> of the kml file. A change in orientation OT (the angle required to go from the current orientation to the desired orientation) is obtained by calculating the difference between OD and OC.

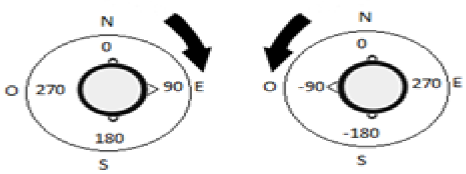


Figure 5. Change of orientation.

$$OT = OD - OC$$

Since

$$0 <= OD < 360$$

and

$$0 <= OC < 360$$

we therefore have :

$$-360 < OT < 360$$

Thus, two cases can be illustrated:

(a) When the user has to move clockwise,

$$(0 <= OT < 360)$$

(b) When the user has to move anti-clockwise.

$$(-360 < OT < 0)$$

Both cases are illustrated in figure 5.

V. EXPERIMENTAL STUDY

There are several systems with which we can compare our system described above. We omit the work presented by Lin et al. [12], since their system indicates only two directions (left and right). We chose to compare our system with standard Google Maps Audio (pedestrian version) and with Pocket navigator [6]. Both applications were installed in Samsung S2 equipped with Android version 2.3. Google Maps Audio is a very popular application. Pocket Navigator is a vibro-tactile system for pedestrian navigation whose principle is described in the first Section of this paper.

Figure 6. Experiment with HaptiNav

A. Protocol

In the experimental study, we carry out three experiments. We vary the navigation mode between the three experiments. The first consists of navigating in audio mode using Google Maps Audio. The second experiment consists of navigating with our own system in a vibro-tactile mode. The third one consists of navigating with Pocket navigator in a vibro-tactile mode. We keep the same route (figure 7) in the three experiments, since each experiment has 12 different participants. The experiments take place at the School of Advanced Industrial Technologies located at Izarbel Science Park. Each subject takes part only in one of the three experiments and is followed by two experimenters. One experimenter managed the dashboard, the other managed the stopwatch. Subjects were not allowed to look at the Smartphone's screen or ask the experimenters questions. They had to walk at their usual speed which is about 1 meter per second. There was no learning phase prior to the experiments.

B. Participants

36 unpaid subjects, 18 female and 18 male, take part in the experimental study. All are ESTIA students, trainees or employees. They are aged between 22 and 39 years old (average =30.5). The 36 subjects are divided into three groups, each with 12 subjects. The first, second and third group take part respectively in the first, second and third experiment. It is worth noting that none of the subjects is involved in the research work presented in this paper. All the users have already used a map and 35 out of 36 are used to using electronic navigation systems, such as Tom-tom. Two subjects are unfamiliar with the experiment's location; the others have already been there. However, knowing the location was not a significant factor since subjects only find out the route to destination at the end of the experiment.

Figure 7. Izarbel Science Park route.

C. *Quantitative study*

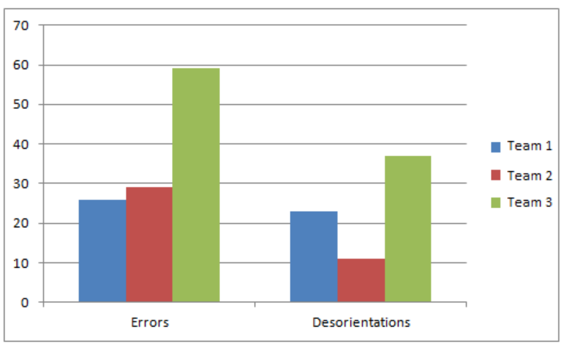


Figure 8. errors and disorientations rates.

The results of the experimental study show that the three techniques can be efficiently used for navigating. Figure 8 shows the number of navigation errors as well as the number of disorientation events. An error is recorded when subjects did not take the right direction with regard to a way point (different from that given by the navigation system). A way point is schematized in Figure 7 by a blue arrow.

A disorientation event is noted when a subject stops or deviates from the simplified route in figure 6 for more than 10 seconds. A disorientation event is also noted when a subject indicates to the experimenter that he is confused. After a navigation error or a disorientation event, the subject is redirected towards the right direction. HaptiNav and Google Maps Audio have a very similar number of errors, 29 and 27 respectively. However, disorientations with Google Maps Audio occur twice as often as disorientations with HaptiNav. They happen with Pocket Navigator three times as often as with HaptiNav. Navigation errors of Pocket Navigator occur more than twice as often as navigation errors of HaptiNav and Google Maps audio. Navigation errors with HaptiNav and Pocket Navigator, take place when the vibro-tactile signal is not understood or due to errors relating to the GPS or digital compass. Navigation errors take place with Google Maps Audio due to GPS errors. Figure 7 shows the number of navigation errors and the number of disorientation events related to the three studied systems.

Figure 9 shows the comparison of averages between the groups in terms of errors. The smallest error recorded for HaptiNav is 0. However, the smallest error recorded for Pocket Navigator is 4.

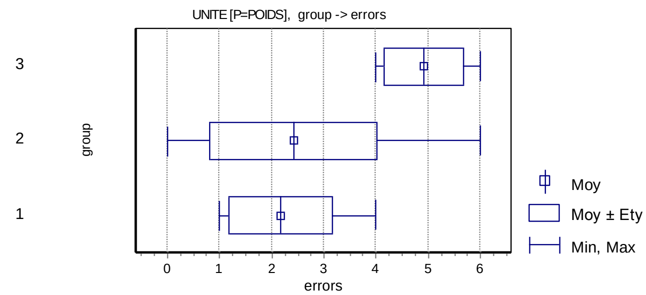


Figure 9. Diagram of averages relating to errors.

Since the ANOVA test gives a value of p equal to 0.6641 ($p=0.6641>0.5$), we conclude that there is no difference between the errors averages of the first and the second experiment. We establish the fact that the performances of HaptiNav are close to those of Google Maps Audio. The ANOVA test gives a p value equal to 0.0001($p=0.0001<0.5$) for the second and third experiment. We therefore conclude that HaptiNav is better than Pocket Navigator in terms of navigation errors.

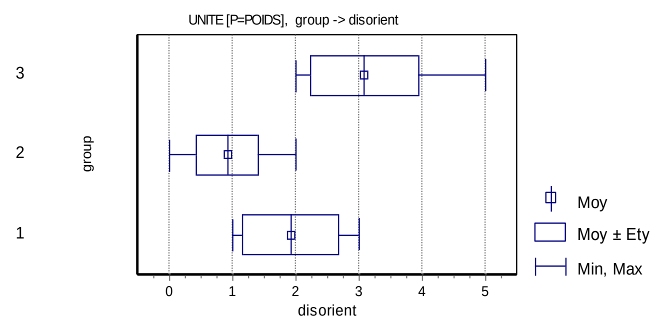


Figure 10. Diagram showing average number of disorientations.

Figure 10 shows the comparison of averages between the groups in terms of disorientation events. Since the value of p ($p = 0.0014 < 0.05$) is below the significance level of 0.05, there is therefore a significant difference between HaptiNav and Google Maps Audio. This enables us to confirm the hypothesis that HaptiNav is more efficient than Google Maps Audio in terms of reducing disorientations during navigation. The ANOVA test highlights a significant difference between Google Maps Audio and Pocket Navigator ($p = 0.0028 < 0.05$). Hence, the PocketNavigator system generates more disorientations events than Google Maps Audio.

Since the value of p ($p = 0.0014 < 0.05$) is below the significance level of 0.05, there is therefore a significant difference between HaptiNav and Google Maps Audio. This enables us to confirm the hypothesis that HaptiNav is more efficient than Google Maps Audio in terms of reducing disorientations during navigation.

ISBN: 978-1-61208-250-9

D. Qualitative study

Subjective estimation of workload is measured using the Nasa TLX Load Index test. This is a multidimensional evaluation test in the form of weightings and it is applied to the measurement of six specific load factors: mental effort, time spent, frustration, physical load, performance and effort. Subjects note the system studied by attributing a mark from 0 to 20 for each factor. No significant difference is detected between HaptiNav and Google Maps Audio in terms of physical load, time spent or performance ($p = 0.68 > 0.05$). However, statistical analysis of perceived mental load, frustration and effort dedicated to the task, highlights the advantage of HaptiNav ($p < 0.05$) with regard to Google Maps Audio and Pocket Navigator. These results show that the mental effort perceived by subjects to understand HaptiNav's vibro-tactile instructions, is on average equal to 4.5 points, which did not affect the frustration felt with this system. However, Pocket Navigator is rated with the highest level of frustration. The average of frustration with this system is 17.84 points.

We ask the following question to the second group of subjects (subjects who take part in the navigation experiment with HaptiNav): "Would you accept wearing this belt to find your way around a town which you are visiting for the first time?" 41 % say they would not, which means that 5 of the 12 subjects questioned would refuse to wear the HaptiNav tactile belt system when visiting a town for the first time. These participants explain that they do not like to wear a belt in town because they do not want to be noticed by other pedestrians. They suggest that this system could be smaller. We ask the following question to the third group of subjects (subjects who take part in the navigation experiment with Pocket Navigator): "Would you accept using Pocket Navigator to find your way around a town which you are visiting for the first time?" More than 80% say that they would not because Pocket Navigator is very inaccurate. We ask this question to the first group of subjects (subjects who take part in the navigation experiment with Google Maps Audio): "Would you accept using Google Maps Audio to find your way around a town which you are visiting for the first time?" More than 80% said that they would only if Google Maps Audio is turned in graphic mode in addition to audio mode.

E. Discussion

The obtained results confirm that the HaptiNav system can be used for hands free and eyes free navigation. We consider improving the quality and the intensity of vibrations in the second prototype of HaptiNav system. Indeed, the participants notice that vibrations became difficult to distinguish when the system's belt is worn on top of thick clothes. This issue explains some of navigation errors happened with HaptiNav.

Some participants say that they refuse to wear HaptiNav because of the belt. So, we will replace the belt with a bracelet which can be worn around the wrist [14], in our System's second prototype.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have presented an haptic system "HaptiNav" which can be used to guide users to their destination.The aim of our research is to determine whether augmented reality in haptic modality can be used to ensure hands-free and eyes-free navigation in an urban environment. In order to answer this question, we test our system with people with normal vision. We compare it to Google Maps Audio and to Pocket Navigator systems, to evaluate its performance.

With HaptiNav, all subjects manage to reach their destinations. HaptiNav and Google Maps Audio systems have approximately the same error rate. However, HaptiNav has the advantage of reducing the number of disorientations. In addition, a statistical analysis of the mental workload, frustration and effort highlights the advantage of HaptiNav, compared to Google Maps Audio and Pocket Navigator. These results show the performance of HaptiNav. Some of participants said that they refuse to wear HaptiNav because of the belt. To overcome this problem, we will replace the belt with a bracelet which can be worn around the wrist, in our System's second prototype. We envisage developing our system, so that the new prototype can support navigation by the visually impaired, in an urban environment. We intend to add a proximity sensor to this system, in order to detect obstacles along the route for the visually impaired. We also plan to add a movement sensor to reintroduce the perception of movement lost by visually impaired people.

ACKNOWLEDGMENTS

We would like to thank Mr. Octavian Curea, ENERGEA team director at ESTIA Research and Mr. Guillaume Terrasson, member of the Mechatronics team. We would also like to thank Julien Conon and Simon Garde two students of ESTIA.

REFERENCES

[1] http://arduino.cc/en/main/arduinoboardbluetooth/. February 2012.

[2] http://www.amarino-toolkit.net/. February 2012.

[3] http://www.androidguys.com/2010/04/19/pocketnavigator-tactilepedestrian-navigation/. February 2012.

[4] P. Fuchs, G. Moreau, S. Donikian, 38 auteurs, Le traité de la réalité virtuelle, troisième édition, cinquième volume : « Les humains virtuels ». Les Presses de l'Ecole Mines ParisTech. ISBN 978-2-911256-00-4, March 2009.

[5] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. Mac- Intyre. Recent advances in augmented reality. IEEE Comput.Graph.Appl. Volume.21, pages : 34–47, November 2001.

[6] B. Bayart. Réalité augmentée haptique : théorie et applications. PhD thesis, Robotique, Université d'Evry, December 2007.

[7] S.A. Brewster and L.M. Brown. Tactons: structured tactile messages for non-visual information display. In Proceedings of the fifth conference on Australasian user interface. Volume .28, pages : 15–23, January 2004.

[8] L.M. Brown, S.A. Brewster, and H.C. Purchase. A first investigation into the effectiveness of tactons. In Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. Pages : 167–176, March 2005.

[9] M.S. Jin and J.I. Park. Interactive Mobile Augmented Reality system using a vibro-tactile pad. In IEEE International Symposium on VR Innovation. Pages: 329 - 330 , March 2011.

[10] A.S.G. Panagiotis, D. Ritsos and D.P. Ritsos. Standards for augmented reality: a user experience. International AR Standards Meeting. Pages: 1-9, February 2011.

[11] M. Pielot, B. Poppinga, and S. Boll. Pocketnavigator: vibro- tactile waypoint navigation for everyday mobile devices. In Proceedings of the 12th international conference on Human Computer Interaction with mobile devices and Services. Pages: 423–426, September 2010.

[12] M.W. Lin, Y.M. Chang, W. Yu, and F.E. Sandnes. Investigation into the feasibility of using tactons to provide navigation cues in pedestrian situations. In proceeding of the 20th Australasian Computer-Human Interaction Conference, OZCHI 2008: Designing for Habitus and Habitat. Pages: 299-302, December 2008.

[13] B. Shneiderman. Designing the User Interface: Strategies for Effective Human-Computer Interaction. Addison-Wesley Longman Publishing Co., Inc. ISBN:0-201-16505-8, 1986.

[14] S. Kammoun, C. Jouffrais, T. Guerreiro, H. Nicolau, H. Jorge and J. Guiding. Guiding blind people with haptic feedback. In Pervasive 2012 Workshop on Frontiers in Accessibility for Pervasive Computing. June 2012.

[15] G.H. Yang, M.S. Jin, Y. Jin, and S. Kang. T-mobile: Vibrotactile display pad with spatial and directional information for hand-held device. The 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. Pages: 5245–5250, October 2010.

[16] T. Yliopisto, G. Evreinov, L. Vesterinen, A. Nyman, J. Kminen, J. Jokinen, and D. Mathew. vsmileys: Imaging emotions through vibration patterns. AlternativeAccess: Feelings Games. Pages: 75–80, May 2005.

Haptic Mouse

Enabling Near Surface Haptics in Pointing Interfaces

Kasun Karunanayaka, Sanath Siriwardana, Chamari Edirisinghe, Ryohei Nakatsu
KEIO NUS Cute Center, Interactive and Digital Media Institute, National University of Singapore, Singapore
e-mail: {g0800474,sanath,chamari,elenr}@nus.edu.sg

Ponnampalam Gopalakrishnakone
Yong Loo Lin School of Medicine, National University of Singapore, Singapore
e-mail: antgopal@nus.edu.sg

***Abstract*— In this study, we are introducing an innovative pointing interface for computers, which provides mouse functionalities with near surface haptic. It could also be configured as a haptic display, where users can feel the basic geometrical shapes in the GUI by moving the finger on top of the device surface. These functionalities are attained by tracking 3D position of a neodymium magnet, using Hall Effect sensors grid and generating like polarity haptic feedback using an electromagnet array. Where previously haptic sensations are felt only on top of the buttons of the haptic mouse implementations, this interface brings the haptic sensations to the 3D space.**

Keywords- pointing interface; near surface haptic feedback; tactile display; tangible user interface.

I. INTRODUCTION

Ordinarily, Pointing devices are used to control and provide data to the graphical user interfaces (GUI) using physical gestures [1]. Movements and commands sent by pointing devices are echoed on the screen through movements of the mouse pointer (or cursor) and other visual changes. Mouse is the most common and popular pointing device in use nowadays. Pointing interfaces have been in the use along with computers for almost four decades [9]. They were continuously improved by adding new features like dragging, scrolling, multi-touch and recently, attempts were made to include haptic feedback sensations. It was argued that the addition of haptic sensations will create excitement, realism, and an added natural feel for the users [10].

Haptic Mouse (Fig. 1) is introduced as a new type of pointing interface which will provide mouse interactions, haptic feedback and additional enhanced features. The key advantage of this system over other haptic pointing interfaces is that users are able to control the mouse cursor and feel haptic sensations from 4cm above the device surface. This will enable the haptic sensations in 3D space which will be a novel experience. Varied haptic sensations provided by this system can be felt like attraction, repulsion, and various patterns of vibrations. Those sensations can be easily configurable as feedbacks for different mouse commands using the driver software that we have developed.

The system provides attraction and repulsion sensations by changing the polarity of the electromagnets. Polarity of an electromagnet can be changed by swapping the positive and negative voltage supply to electromagnet using a controller circuit. When the neodymium magnet worn on the finger tips and the electromagnet array underneath the device positioned in the opposite polarity, (N – S or S - N) users feel an attraction towards the device surface. When those magnets are in like polarity (S - S or N-N) positions users feel the repulsion sensation.

Vibration sensations are initiated by setting up neodymium magnet and magnetic array in a like polarity position and then rapidly switching on and off the electromagnets in the array in certain frequencies. This rapid switching on and off dynamically changes the magnetic field it produces and affects the static magnetic flux developed by the neodymium magnet worn on the finger tips. While electromagnet is switched off, neodymium magnet drop towards the device surface, but when the electromagnet is switched on, the neodymium magnet rises due to the repulsion force felt as a vibration.

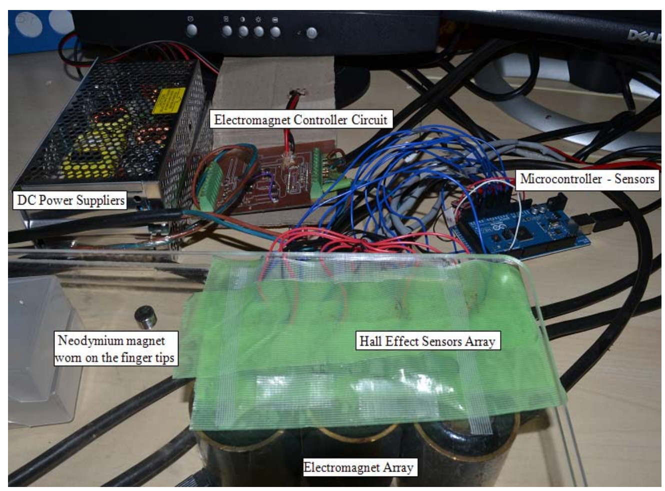
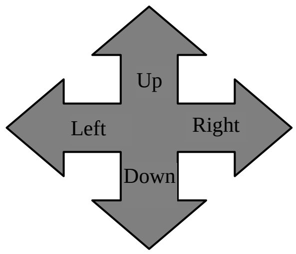
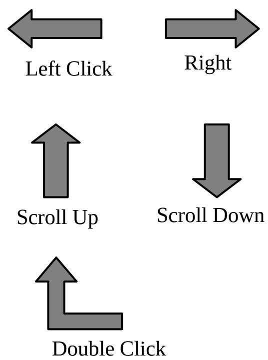


Figure 1. Haptic mouse system : User can move the neodymium magnet worn on the finger tip above the device surface and interact with the computer & sense haptic feedback for their inputs.

Haptic Mouse allows users to both control and interacts with the graphical user interfaces as similar to other pointing interfaces. Figure 2, shows the related gestures of 2D cursor movements and Commands of the system. Users can easily move the neodymium magnet attached to the finger tip on top of the device surface and control the GUIs. This can be visualized as a movement of an invisible mouse on the mouse pad.



2a. Haptic Mouse cursor movement gestures while North Pole is positioned downwards



2b. Haptic Mouse commands while South Pole is positioned downwards

Figure 2. movement and Command gestures of the Haptic Mouse

Furthermore, this interface can be configured as a low resolution haptic display. It is possible for a user to move his/her finger on top of the surface and sense the basic shapes of the objects on the screen. This is achieved by providing one vibration pattern once the user moves the cursor on top of the interested object and then change to a different vibration patterns when the cursor crosses the border of that object. Simple geometrical shapes which are bigger than 200 pixels * 200 pixels can be sensed and identified. As further developments, if there is an application which is restricting the user to a particular window, this device can use the haptic feedback and let the user know about the virtual boundary. The sensing of simple gestures will be helpful for users to increase their interaction with computers.

This paper will discuss the prior research and publications as related work, where the limitations and research gaps will be highlighted. The implementation details of the Haptic Mouse will be discussed extensively in the system description section while the results section will be presenting two technical experiments. The Conclusion & Future Work section presents the end users perspectives and the potential directions for further research.

II. RELATED WORKS

This section will discuss prior research with which the authors are arguing for the novelty value of the Haptic Mouse.

Liquid Interface [2] is a previous work of the authors which has provided the base technologies for the current pointing interface. It is an organic user interface that utilizes ferrofluid as an output display and input buttons embodied with musical notes. Using a matrix of Hall Effect sensors, magnetic fields generated by neodymium magnets worn on the fingertips are measured and then converted into signals that provide input capability. This input actuates an array of electromagnets and generates ferrofluid bubbles. By matching like polarities between the electromagnets and the neodymium magnet, haptic force feedback can be achieved. This system is limited to detect switch on and switch off type of interactions and used to develop a ferrofluid based piano.

FingerFlux [4] is an output technique which generates near-surface haptic feedback on interactive tabletops. It combines electromagnetic actuation with a permanent magnet attached to the user's hand. FingerFlux provides enhanced features like, feel the interface before touching, attraction and repulsion, development of applications such as reducing drifting, adding physical constraints to virtual controls, and guiding the user without visual output. They have achieved the vibration sensations up to 35mm above the table. As limitations, Fingerflux could only works with table top computers. It uses camera tracking based sensing and the maximum vibration feeling height is comparatively lower than our system.

Tactile Explorer [5] is a device which provides access to computer information for the visually disabled based on a tactile mouse. The tactile mouse resembles a regular computer mouse, but differs in having two tactile pads on top that have pins that move up and down. These translate the data on the screen to tactile sensation. Tactile Explorer provides possibilities to find and select desirable on-screen information and study it with different options.

Microsoft tactile mouse [6] is a commercially available mouse implementation which combines haptic sensation and will be developed to support rich features of their latest operating system. This mouse has a touch sensitive strip which contains two buttons, one on each end. Haptic-feedback, in the form of vibration through the touch-sensitive strip, indicates which one of the three scrolling speeds has been selected. Both Tactile Explorer and Microsoft tactile mouse are mouse implementations combined with Haptic. It supports enhanced haptic interactions. However, operations and sensations are limited to the device surface. Furthermore, the haptic actuation is limited to a small area of the device surface.

III. SYSTEM DESCRIPTION

Haptic Mouse contains three modules. They are Neodymium Magnet attached to the finger and Hall Effect sensors Grid, Software Interface Driver and Electromagnetic Array with Controller Circuit. These three parts are described in the following sections.

A. Neodymium Magnet and Hall Effect sensors Grid

The neodymium magnet attached to the finger tip allows users to actuate the Hall Effect sensors grid which is placed bellow the acrylic surface. Polarity of the neodymium magnet and various gestures made by the user are identified with the help of Hall Effect sensors grid. Neodymium magnet could generate higher density of magnetic flux compared to other permanent magnets. Therefore, the size and weight of the permanent magnet is important for this system to become smaller in order to fit it on the fingertip of the user.

We have used an Arduino [3] based microcontroller for processing the Hall Effect sensor readings. Analogue voltage readings of the sensors then converted to digital values using the built-in analog to digital converters and fed in to interface driver software to identify the gestures and commands. Hall Effect sensor grid used in this device is a 4*3 array (4 sensors along the X axis and 3 sensors along the Y axis). The space between two Hall Effect sensors were allocated 100 pixels. In physically, distance betttween two sensors is 2cm for X axis and 1.36cm for Y axis. All the sensor values recorded are represented as X,Y coordinates (0-300 in X axis and 0 to 132 in Y axis).

B. Software Interface Driver

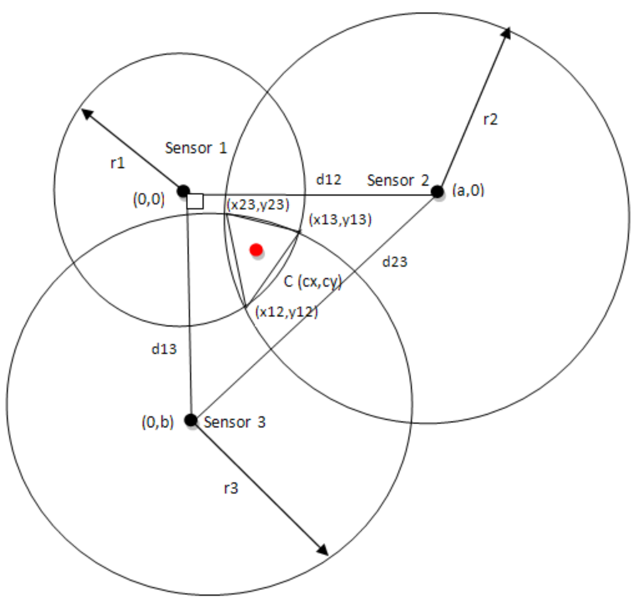


Figure 3. Based on the distance to the neodymium magnet from each sensor the position of the neodymium magnet is determined.

For the precise operation of the pointing device, there has to be a device driver which can integrate with the operating system. Therefore, using the Windows API we have developed a software driver for this device. This driver accepts the digital sensor readings from the microcontroller of the Hall Effects sensors grid as the input. When the North Pole of the neodymium magnet is positioned downward, sensor values are fluctuating in between 512 -1024 range and when the South Pole is downward, sensor values are changing between 0 and 512 of range. These sensor values are sorted in the descending order and if the magnet is North Pole downwards, software searches for the sensors in the grid where it received the maximum readings. Sensors which are nearest to the neodymium magnet, output the maximum values in this case. Based on those intensity values, relative distance to the neodymium magnet from the nearest three sensors is calculated.

The localization algorithm of the neodymium magnet is derived from the 2D trilateration technique. Trilateration is the process of determination absolute or relative locations of points by measurement of distances using the geometry of circles, spheres or triangles [13]. It is widely used for localization algorithms which can be found in GPS, RF based indoor positioning, navigation, and survey systems. As illustrated in Figure 3, the distances to the neodymium magnet is calculated from the expressions found in the table 1. These distances can form circles and based on their intersection position of the sensor can be located. To increase the probability for intersections we have multiplied the distances by a constant factor (which is greater than one) and made the three circles intersect most of the times. This forms a circular triangle [(x_{12}, y_{12}), (x_{13}, y_{13}), (x_{23}, y_{23})] and by finding the center the position of the neodymium magnet can be obtained. The mathematical method we derived is explained bellow.

Let us assume that all there circles formed by the readings are intersect with the each other. Therefore, the conditions for the circles to intersect are as follows.

$$r_1 - r_2 < d_{12} < r_1 + r_2 \qquad (1)$$

$$r_2 - r_3 < d_{23} < r_2 + r_3 \qquad (2)$$

$$r_1 - r_3 < d_{13} < r_1 + r_3 \qquad (3)$$

By following Fewell's [12] method to calculate area of circular triangles, we can calculate the coordinates of the intersection points, (x_{12}, y_{12}), (x_{13}, y_{13}) and (x_{23}, y_{23}) as follows. By defining the origin of the coordinate system is placed at circle 1 and X axis is passed through the center of the circle 2,

$$x_{12} = \frac{r_1^2 - r_2^2 + d_{12}^2}{2d_{12}} \qquad (4)$$

$$y_{12} = \frac{1}{2d_{12}}\sqrt{2d_{12}^2(r_1^2 + r_2^2) - (r_1^2 - r_2^2)^2 - d_{12}^4} \qquad (5)$$

By assuming origin of the (x',y') system is located at the center of the circle 1 and x' axis passes through center of circle 3,

$$x'_{13} = \frac{r_1^2 - r_3^2 + d_{13}^2}{2d_{13}} \qquad (6)$$

$$y'_{13} = \frac{-1}{2d_{13}}\sqrt{2d_{13}^2(r_1^2 + r_3^2) - (r_1^2 - r_3^2)^2 - d_{13}^4} \qquad (7)$$

By transform back to the (x,y) coordinates system and obtaining (x_{13}, y_{13}),

$$x_{13} = x'_{13}\cos\theta' - y'_{13}\sin\theta' \qquad (8)$$

$$y_{13} = x'_{13}\sin\theta' + y'_{13}\cos\theta' \qquad (9)$$

$$\sin\theta' = \sqrt{1 - \cos^2\theta'} \qquad (10)$$

$$\cos\theta' = \frac{d_{12}^2 + d_{13}^2 - d_{233}^2}{2d_{12}d_{23}} \qquad (11)$$

By assuming origin of the (x'',y'') system is located at the center of the circle 2 and x'' axis passes through center of circle 3,

$$x''_{23} = \frac{r_2^2 - r_3^2 + d_{23}^2}{2d_{23}} \quad (12)$$

$$y''_{23} = \frac{1}{2d_{23}}\sqrt{2d_{23}^2(r_2^2 + r_3^2) - (r_2^2 - r_3^2)^2 - d_{23}^4} \quad (13)$$

By transform back to the (x,y) coordinates system and obtaining (x_{23}, y_{23}),

$$x_{23} = x''_{23}cos\theta'' - y''_{23}sin\theta'' + d_{12} \quad (14)$$

$$y_{23} = x''_{23}sin\theta'' + y''_{23}cos\theta'' \quad (15)$$

$$sin\theta'' = \sqrt{1 - cos^2\theta''} \quad (16)$$

$$cos\theta'' = -\frac{d_{12}^2 + d_{23}^2 - d_{13}^2}{2d_{12}d_{23}} \quad 17)$$

θ' and θ'' are the angles between the x axis and respective abscissas of the two additional coordinate systems. After calculating the intersections of the circular traingle we can obtain the center of the traingle where the position of the neodymium magnet exsists is as follows.

$$C_x = \frac{(x_{12} + x_{13} + x_{23})}{3} \quad (18)$$

$$C_y = \frac{(y_{12} + y_{13} + y_{23})}{3} \quad (19)$$

By finding the position of the neodymium magnet and comparing it with the next position, relative X,Y displacement can be calculated. Then these relative displacements are mapped to the last coordinates of the mouse curser position and moves the cursor to a new X,Y location.

In the case of identifying mouse commands, firstly, driver identifies the neodymium magnet which is placed South Pole downwards by reading the digitally converted values. If the magnet is South Pole downwards, software driver searches for the three minimum sensor reading values and determines the coordinates of those sensors. Then, the distance to the neodymium magnet from each sensor is calculated and its position is determined. The movement path of the neodymium magnet is tracked and if the path follows the gestures defined for the mouse commands, the driver activates the appropriate commands. As the final step, it updates Electromagnet controller circuit about the necessary vibration pattern which would eventually provide the user with the vibration feeling.

In the case of sensing the shapes, the driver software keeps a selected vibration pattern until the user move the mouse cursor on top of the interested object in the screen. Once the cursor is moved away from the object boundary, driver sends commands to the microcontroller of the electromagnet controller circuit to change the output frequency.

The sensors were capable to detect the the position accuratly more than 90% of points. Mouse curser position is only when there are two ajecent accurate neodymium magnet position readings exsists and then added the difference of the X,Y displacement to the previous cursor position. Therefore, the accuracy of the movement of the mouse cursor was improved. When the neodymium magnet is placed from 2 cm above the device surface, sensors were only capable to track position about 60% of the movements. In 4cm of hight above the device surface, the sensor array only able to track the position of the neodymium magnet less than 40% of the movements. In this hight, the curser movement became fairly difficult. Because interface driver have to cancell out inaccurate localization data and use only the accurate data to move the mouse pointer

C. *Electromagnetic Array and Controller Circuit*

This part of the system is made with six electromagnets, Magnet controller circuit and Arduino based microcontroller. As the current required by the six electromagnets is 13A [7], it becomes necessary to control the power supplied to the electromagnets via a relay circuit. This is because the voltage and current from the microcontroller pins amounts is only 5V, and 40mA respectively [3], which is insufficient to drive the electromagnet. To address this, the relay circuit acts as a mechanism that is able to switch on a much larger power to drive the electromagnets. For this power up electromagnets, six N-Type MOSFET [8] were used, one for each electromagnet. The connections to the MOSFET are configured such that the MOSFET will enter linear region and produce a drain current ID, of approximately 1.9A when the Arduino outputs a 5V signal to turn on the electromagnet. When the Arduino outputs 0V signal and the MOSFET turns off, the drain current drops to 0A which turns off the electromagnet as illustrated in the figure 4.

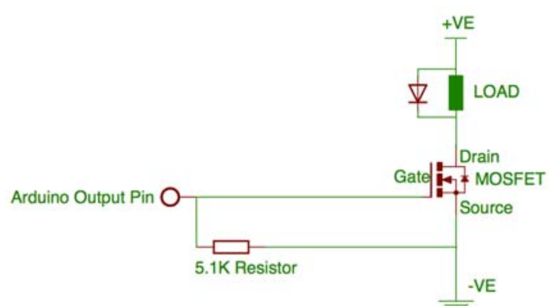


Figure 4. MOSFET based magnet driver circuit

By programming the Arduino to switch continuously from high to low and vice versa in rapid succession, a PWM output wave is produced which in turn causes the MOSFET to continuously turn on and off the main power supply like a relay, generating another PWM output signal with enough current to drive the electromagnet. A diode connected in parallel to the electromagnet to prevent the damage to MOSFET by the backflow of current. A resistor is connected in parallel to the Arduino Pin and acts as a safety turn off mechanism. This design was replicated 6 times to drive the 6 electromagnets.

The electromagnets require PWM to run. The purpose of PWM is to simulate an analog voltage by rapidly toggling a digital pin between on and off. Software Interface Driver sends a 20 character length data frame for every 10ms via the serial connection to the microcontroller to activate the required electromagnets. These data frames are interpreted as commands to turn on the electromagnets that correspond to the haptic feedback sensations felt by the user. Due to the limitations of the electromagnet, the maximum frequency that can be achieved is 100 Hz. Therefore, different frequencies between 5 Hz to 100Hz were used to provide different Haptic sensations to the users.

IV. Results

Two technical experiments were carried out to measure the capabilities and limitations of the system.

A. Hall Effect sensor reading versus horizontal and perpendicular distances to the neodymium magnet

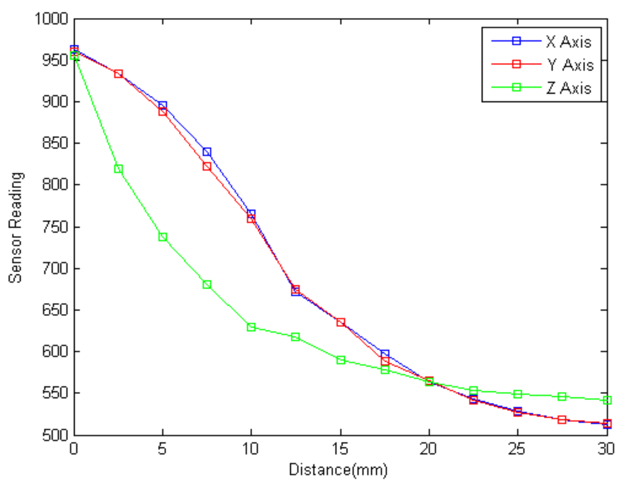


Figure 5. Plot of Sensor Output vs. perpendicular distance to the neodymium magnet

TABLE I. Exprissions to determine the distance between the Hall Effect sensor and neodymium magnet based on the Hall Effect readings

Distance (mm)	**Deriving distance calculation expressions**		
	M	***C***	***Expression***
0 - 2.5	-12	963	Y = -12X + 963
2.5 – 5	-15.2	973	Y = -15X + 971
5 – 7.5	-22	1005	Y = -22X + 1005
7.5 – 10	-30	1065	Y = -30 + 1065
10 – 12.5	-37.2	1137	Y = -37.2X + 1137
12.5 – 15	-14.8	857	Y = -14X + 857
15 – 17.5	-15.2	863	Y = -15.2X + 863
17.5 – 20	-13.6	835	Y = -13.6 + 835
10 – 22.5	-8	723	Y = -8X + 723
22.5 – 25	-6	678	Y = -6X + 678
25 – 27.5	-3.6	618	Y = -3.6X + 618
27.5 30	-2.4	585	Y = -2.4X + 585

The objective of this experiment is to investigate the variation in the magnetic field strength vs. the distance of all thee axis and determine the strength of the magnetic field needed to be produced by the neodymium magnet to achieve the desired tracking ability. The experiment was conducted by positioning the neodymium magnet on top of the Hall Effect sensor and measure output readings at various distances in all three axis and results are shown in the figure 5. According to the results shown in figure 5, it is clear that sensor reading values are following nonlinear curves but along the X and Y axis the readings are approximately the same. Therefore based on the X and Y axis sensor readings we have derived set of equations to calculate the distance between the sensor and the neodymium magnet which is presented in the Table1. As a result these expressions provide the distance values needed by 2D trilateration based localization algorithm to obtain the position of the neodymium magnet.

B. Height of Haptic sensation is felt vs. Pulse Width Modulation for different voltage levels

The purpose of the experiment is to evaluate the relationship between the PWM and the maximum height that haptic sensations can be sensed above the device surface by running ellectromagnets in three different voltage levels. With the results, the PWM values and voltage levels that correspond to achieve haptic sensations in certain heights can be determined. In addition, the optimal PWM values that need to be set during actuation can be verified.

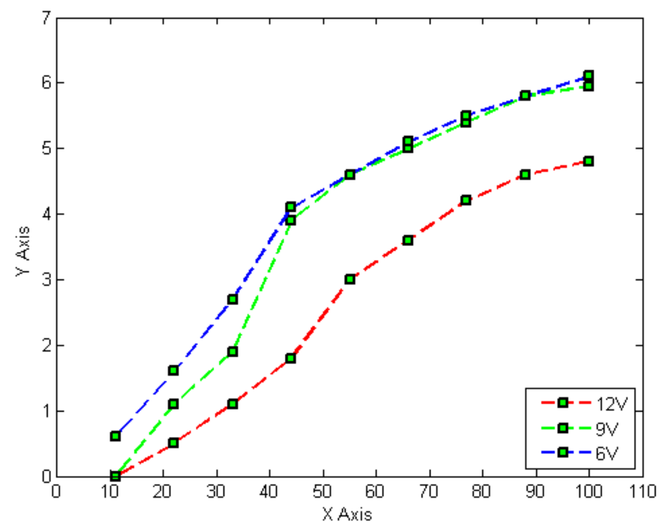


Figure 6. Plot of the maximum haptic sensation height achieved vs. PWM values.

As shown in Figure 6, the relationship between the PWM and the maximum height that haptic sensation can be sensed follows an increasing linear trend in all three different voltage levels, suggesting that the system is linearly controllable. However, it is noted that the haptic sensations were started to feel from PWM running on 11% for 6v and 9v mode but when electromagnets operate on 12v haptic sensations can be felt in 0.6 cm. Further, when the PWM values are between 90% and 100%, it is hard to notice the maximum difference of the actuation for 9v and 12 v. With these results we were able to provide haptic sensations up to 6.1cm above the device surface. However, by changing the type of the electromagnets, it may be posible to increase the height that heptic sensations can be felt.

V. CONCLUSION & FUTURE WORK

To conclude, in this paper we have presented a new type of computer interface which provides basic pointing interface functionalities with near surface haptic feedback up to 6 cm of height. The advantages and limitations of the system were also discussed with the related works. Using two technical studies we were able to show that system can perform in an adequate manner and has strong potentials to be improved. Haptic Mouse provides the base tools to combine magnetic field based devices with computers.

The haptic feedback resolution of the interface can be improved by using smaller electromagnets in larger numbers. When the device is used as a haptic display, an increased resolution would offer a better accuracy and representation of the virtual objects in the computer screen. Moreover, implementing variable friction for haptic interface using this technology will be an interesting research topic in future since variable friction has not been implented for touch sensitive haptic feedback systems.TeslaTouch [11] is the closests execusion of such haptic display .

This device has great potentials to be improved as an interface for visually handicapped who rely mostly on touch sensation. In orderto improve to this level of proficiency, this system is required to minimize the size of the electromagnets and increase the density of electromagnets packed in the electromagnets array which will provide a better resolution. This device could also be improved as an easy learning tool for children, which can be used to draw some basic shapes or characters that will enhance the interactive enjoyment.

ACKNOWLEDGMENT

This research is carried out under CUTE Project No. WBS R-7050000-100-279 partially funded by a grant from the National Research Foundation (NRF) administered by the Media Development Authority (MDA) of Singapore.

REFERENCES

[1] Maria Virvou and Katerina Kabassi. 2002. Reasoning about users'actions in a graphical user interface. Hum.-Comput. Interact. 17, 4 (December 2002), 369-398.

[2] Kasun Karunanayaka, Jerey Tzu Kwan Valino Koh, Eishem Bilal Naik, and Adrian David Cheok, "Hall Effect Sensing Input and Like Polarity Haptic Feedback in the Liquid Interface System", in Proceedings of the international joint conference on Ambient Intelligence, AmI '11. Amsterdam, Netherlands, 2011.

[3] Arduino. Arduino. [Online]. http://www.arduino.cc/en/

[4] Weiss, M., Wacharamanotham, C., Voelker, S., and Borchers, J. 2011. FingerFlux: near-surface haptic feedback on tabletops. In Proceedings of the 24th Annual ACM Symposium on User interface Software and Technology (Santa Barbara, California, USA, October 16 - 19, 2011). UIST '11. ACM, New York, NY, 615-620.

[5] Tactile Explorer, [Online]. http://www.tactile-world.com/

[6] El Saddik, A., Orozco, M., Eid, M., Cha, J., Haptics Technologies Bringing Touch to Multimedia, Series: Springer Series on Touch and Haptic Systems, 2011.

[7] Magnetech Corporation. Magnetech Corporation. [Online]. http://www.magnetechcorp.com/Opposite_Pole.htm

[8] MOSFETs, [Online]. http://www.falstad.com/circuit/e-nmosfet.html

[9] Engelbart, D. C. et al. (1968). A Research Center for Augmenting Human Intellect. (demonstration) Stanford Research Institute, Menlo Park, CA. http://sloan.stanford.edu/MouseSite/1968Demo.html

[10] A. El Saddik, "The potential of haptics technologies," IEEE Instrumentation & Measurement Magn.,vol. 10, pp. 10-17, Feb. 2007.

[11] Bau, O., Poupyrev, I., Israr, A, Harrison, C. TeslaTouch: Electrovibration for Touch Surfaces. Proceedings of UIST 2010: ACM: pp. 283-292.

[12] Fewell, M. "Area of Common Overlap of Three Circles." Australian Dept. Defense. Oct. 2006. http://www.dsto.defence.gov.au/publications/4815/DSTO-TN-0722.pdf. Retrived January 10, 2013

[13] D.E. Manolakis, "Efficient Solution and Performance Analysis of 3-D Position Estimation by Trilateration, " IEEE Trans. on Aerospace and Electronic Systems, Vol. 32, No. 4, October 1996, pp. 1239-1248.

Fundamental Study to Consider for Advanced Interface in Grasping Movement

Shunji Shimizu / Tokyo University of Science, Suwa
Department of Electric Systems Engineering
Chino-city, Japan
shun@rs.suwa.tus.ac.jp

Hiroaki Inoue/ Tokyo University of Science, Suwa
Research course of Engineering/Management
Chino-city, Japan
jgh12701@ed.suwa.tus.ac.jp

Noboru Takahashi/ Tokyo University of Science, Suwa
Research course of Engineering/Management
Chino-city, Japan
srl@rs.suwa.tus.ac.jp

***Abstract*— the analysis of human grasping movement is important in developing methodologies for controlling robots or understanding human motion programs. In analyzing human grasping movement, it is advantageous to classify movements. In previous papers, classifications of grasping patterns were proposed according to the posture. Among these classifications of grasping patterns, no unified view has been reached as yet. The measured quantities in grasping have included only the posture of the hand, force and its distribution. Few have pertained to classifications based on grasping force and its distribution. This paper first tries to analyze the effect of visual information on grasping movements, and then attempts to classify grasping movements broadly according to their purpose. For the elements of the purposes of grasping movements, movements that were decided upon were those which require attention, snapping or the adjustment of the wrist or movements which do not require any special action to achieve their purpose. Secondly, we focus on the tactile information to predict with a limitation of movement. Finally, we attempted to discuss the relation between human brain activity and grasping movement on cognitive tasks.**

Keywords-human hand; grasping force; grasping pattern; brain activity;NIRS.

I. INTRODUCTION

Human hands are so dexterously controlled that they can manipulate almost anything freely. Observations obtained from analyzing the grasping patterns of human hands will be useful in the control of robot hands. For example, it may be possible for industrial robots to deal flexibly with and solve unexpected problems which may occur. In the construction of more sophisticated interface systems the analysis of the grasping patterns of human hands is also suggested as an important subject for controlling robot hands by remotely.

To analyze grasping patterns, many researchers, including Schlesinger [1] have proposed and reported methods to classify grasping patterns [2][3]. However, these classifications of grasping modes depend largely on the researcher's personal definitions, and no unified view has been reached at present.

The measured quantities in grasping include the posture of the hands, and force and its distribution. However, most classifications have been based on the posture of the hands, and little has been reported on classification based on grasping force and its distribution.

From the point of the view of the grasping task, Napier broadly divided grasping patters into "power grip" and "precision grip" [4]. In addition, Cutkosky classified more grasping patterns by incorporating details of the objects and the precision of the task in Napier's concept [5]. Meanwhile, Kamakura et al. presented a classification cased on the contact pattern between the grasped object and grasping hand [6]. Kang et al. proposed the technique of the "contact web" which estimated information and classified grasping based on the resulting contact pattern [7]. For the theory of multi-fingered hands Yokokawa is proposing the dynamic multi-fingered manipulability measurement under the concept of the dynamic manipulability [8]. Another new approach to control the robots is the Programming by Demonstration done (PbD). A late report is proposed by K. Bernard in et al [9]. Shimizu et al. described the Sensor Glove MKIII which is useful in analyzing grasping patterns and shows the potential of measuring grasping force distribution for classification [10].

Many reserch works exist on the classification of grasping movements. However, there are only a few areas using these classifications, such as the industry, indicating a need for a more useful grasping classification to be used in engineering. We thus considered using new elements to broadly classify grasping movements. We set grasping movement purposes for the new elements, elements which are movements that require attention, snapping or adjustment of wrist or do not require any special action to achieve their purpose. Therefore, it was necessary to find a place in which position and direction had little effect when measuring grasping movements. We subsequently deliberated the possibility of broadly classifying grasping patterns according to the purpose of the grasping movement. A report about the importance of grasping task is proposed by Shiraishi, et al. [11], too.

This paper first tries to analyze the effect of visual information on grasping movements and then considers the potential for using the elements described above in classification. Lastly, we focus on the tactile information to predict with a limitation of movement.

 ISBN: 978-1-61208-250-9

II. EXPERIMENT FOR MEASURING GRASPING MOVEMENT WITH CONSIDERING EFFECT OF WRIST DIRECTION

A. Experimantal Method 1

To measure grasping movement with minimal effect from the position and object's direction for classification, it is necessary to ascertain the proper position and direction from which to do so. The next step is to discover the role visual information plays in grasping patterns. In this experiment, USB cameras from three directions measured grasping patterns. The cameras used had a resolution of over 0.3 megapixels.

B. Range of Movement

To find the area which would not need to be considered in regard to its effect on grasping shape; subjects were made to grasp a pointer directed toward them. Then the area in which they could move without changing the direction was measured. Fig. 1 shows the range of movement. Subjects were four healthy, right-handed men aged 22 to 24.

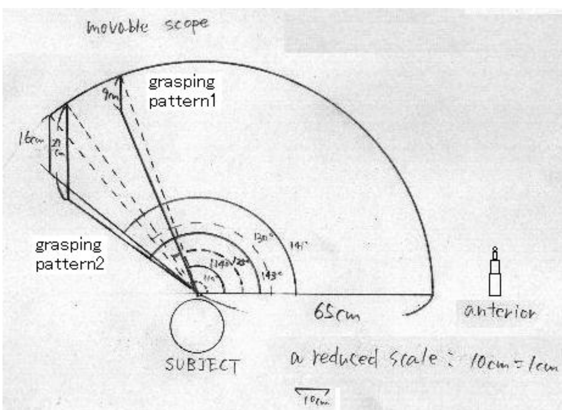


Figure 1. Range of movement (object: pointer)

The range of movement seems to depend on their flexibility and grasping patterns. As can be seen in Fig. 1, a 15 cm margin has been set, and the objects have been placed to check their effect on grasping. Fig. 2 shows the position in which the objects were placed. Grasping patterns differed little according to position; here, Position C was used for classification. However, if the object is placed in direction (4), the grasping patterns change for right-handed people. The next step was to look at the relationship between direction and grasping

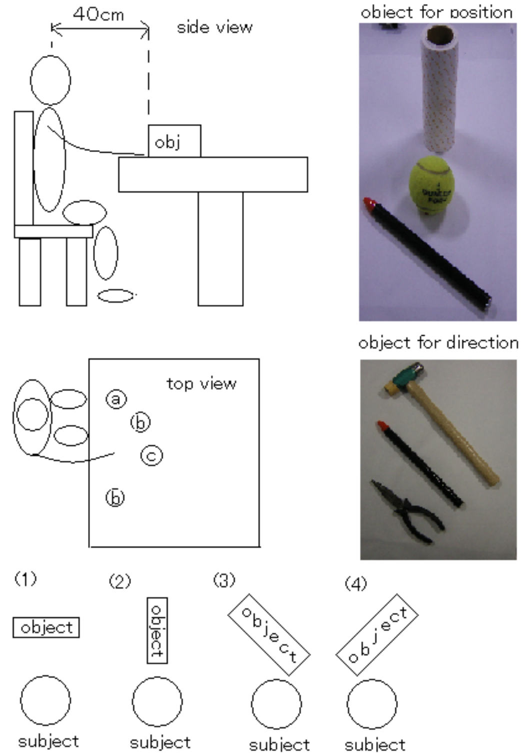


Figure 2. object's position and direction

C. Effect of Direction on Grasping

To analyze the relationship between direction and grasping pattern, the angle of the underarm and the angles in Fig. 3 were measured by changing the object's direction. The directions of (1) to (3) were measured by changing the angle 30 degrees. Subjects were told to grasp the object and put it onto another table. The subjects were five right-handed healthy men aged 22 – 24.

Increases in the angle of the wrist and the angle of the finger baseline are seen to be related to the angle of the object. The angle of the underarm decreased as the angle of the object increased. However, as the angle of the underarm is influenced by reaching, the displacement is not uniform. Fig. 4 exemplifies the difference between grasping patterns and the object's directions.

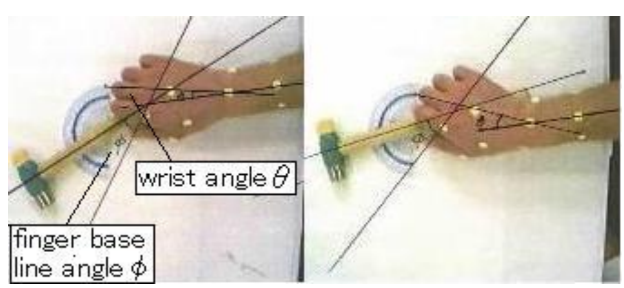


Figure 3. angle measured, θand φ

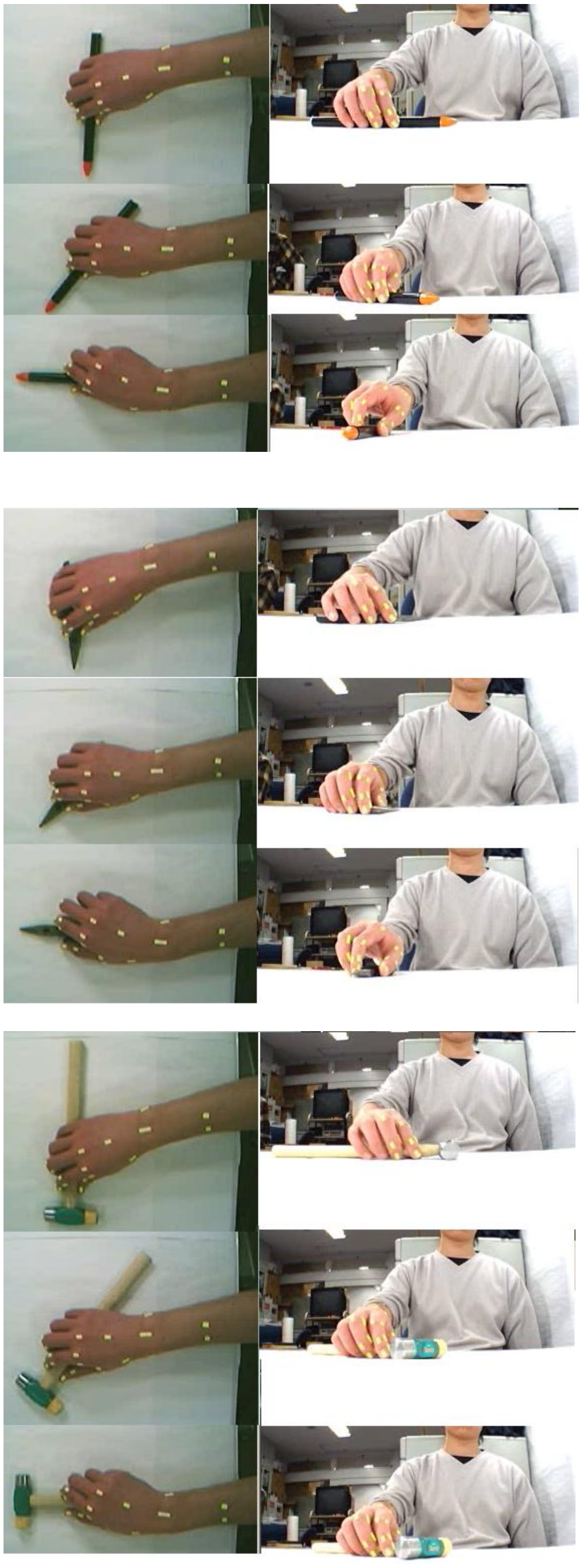


Figure 4. grasping patterns for each object

(Top: pointer, Middle: pincher, Bottom: hammer)

III. Experiment of Classification of grasping movements

A. Experimental Method 2

We compared the difference of grasping movement based on “purpose of grasping movement”. We set elements which are movements that require attention, snapping or adjustment of wrist or do not require any special action to achieve their purpose. We considered the potential for using the elements described above in classification [12].

B. Experiment for the Classification of Grasping Movements

To classify grasping movements, the purpose of grasping was used as an element that is determined at a certain point, i.e., before or after grasping. In this experiment, five kinds of tasks were used to consider the relation between grasping patterns and their purpose. The tasks are to move the object to another place (Task 1), to put the object onto a small box (Task 2), to throw the object (Task 3), to make the object pass through a small hole (Task 4), to use the object as you usually do (Task 5) after grasping. These tasks were created to measure a simple grasping movement (Task 1), a movement requiring attention (Task 2), the movement which requires snapping (Task 3), a movement which requires attention and adjustment of wrist (Task 4), a movement that is imagined to be associated with the object (Task 5). And checked that the grasping movements would change with there tasks.

The purpose of the first experiment was to measure grasping shapes according to their purpose. The next step was to measure grasps without a prescribed purpose. The purpose was given only after the subject first grasped the object. The subjects were six right-handed healthy men aged 22 – 24.

C. Result of Classfication Experiment

The results showed two movements involved in grasping and taking action when checking the difference between the first grasping shape and the next grasping shape. One is a change in grasping shapes before picking up. The other one is a change in the grasping shapes after picking up. Fig 6 shows the first movement; the grasping shape changes according to its purpose. T shows the second movement; the grasping shape changes according to its purpose. Such cases might be difficult to classify only by force distribution. Such movements were not seen in the case of every object. Therefore we attempted to discover out the difference between them and use it with the force distribution to classify grasping movements.

This experiment showed that one seems to make space between the palms and object or change the grasping shape into a more flexible form when they try to do something sensitive like Task 4. Probably, such a tendency has some relation to degree of freedom in movement.

 ISBN: 978-1-61208-250-9

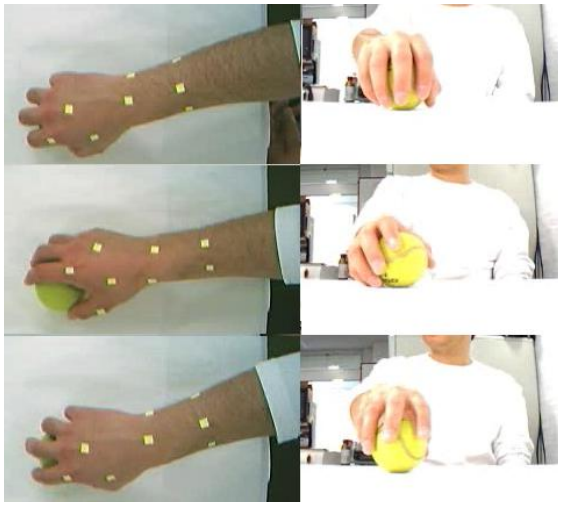


Figure 5. changing the shape when grasping

(Top: Task 1, Middle: Task 3, Bottom: Task 4)

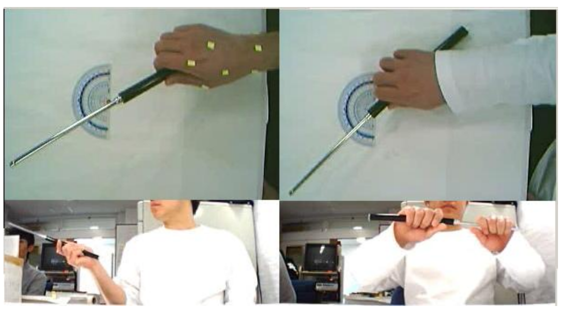


Figure 6. changing the shape after grasping

(left:task4, right:task5)

IV. TACTILE INFORMATION

Looking into the tactile information and degree of freedom in hand movement, we first tried to check how correctly we could imagine the shape only from tactile information. Therefore, the relation of such information to the degree of freedom in movement was considered in conducting this experiment [13].

A. Experiment Method for Tactile Information

Three limitations were made to analyze the effect. One was wearing an eye mask to shut out the visual information. Another one was a pinching movement to control the tactile information. The last one aluminum fingertip cover to reduce the tactile information and to make the surface like robot hand because we are thinking to use the results for robot's hands. The fingertip cover used in this experiment was enclosed in aluminum sheeting and the finger cushion's side was flattened. Limitations in pinching movement are shown in Fig 7.

The test was to guess the object with eyes masked and fingertip limitations. After that, same tests were conducted without using the fingertips. Fig 7, 8 and 9 shows the objects used. Objects in Fig. 8 were stuck to board. The objects in Fig. 8 (lower right) can be spun using the stick that is standing on the small plate. The objects in Fig. 9 can be pinched freely.

TABLE I. LIMITATION OF PINCH

freedom degree number	Limitation
1	Not allowed to pinch again
1'	allowed to grasp again just a little
2	allowed to move up and down
3	allowed to go over lateral side
4	allowed to grasp freely

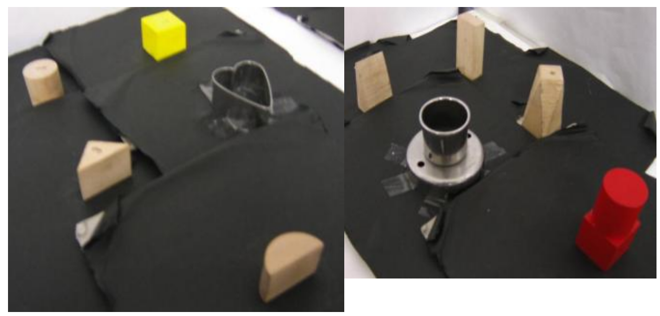


Figure 7. object for tactile information test

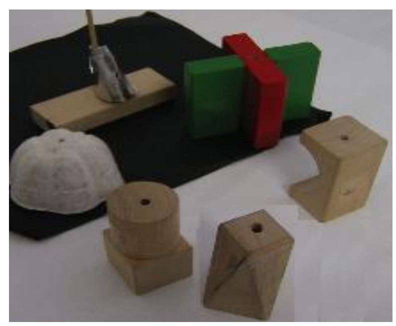


Figure 8. Object for tactile information test

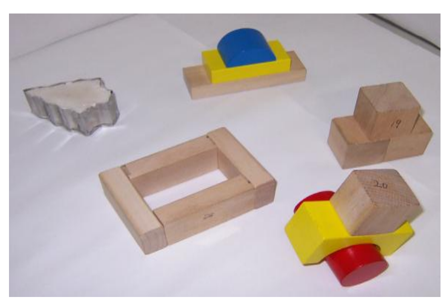


Figure 9. Object for tactile information test

 ISBN: 978-1-61208-250-9

B. Result of the Experiment in Tactile Information

Fig. 10 shows the result of this test. The answers were checked by a majority decision of three observers. The subjects for this experiment were six men and two women, all healthy, right-handed and aged 20 – 60.

In Fig. 10, the accuracy rate was higher when the fingertip covers were not used, and the accuracy rate basically increased with the degree of freedom, but it increased only slightly when the fingertip covers were not used, or when there was a high degree of freedom. From the aspect of object identification only with tactile information, the difference in fingertip cover suggests the importance of the ridges in the fingers' skin and the plasticity of the finger surface, probably because they are enhancing the signals. Furthermore, Degree of Freedom 4 was lower than that of 1 when they tried without the fingertip cover. This seems to have occurred when they lost track of direction when they moved their hands. This suggests that the relation between accuracy rate and freedom digger is not a simple proportional relation.

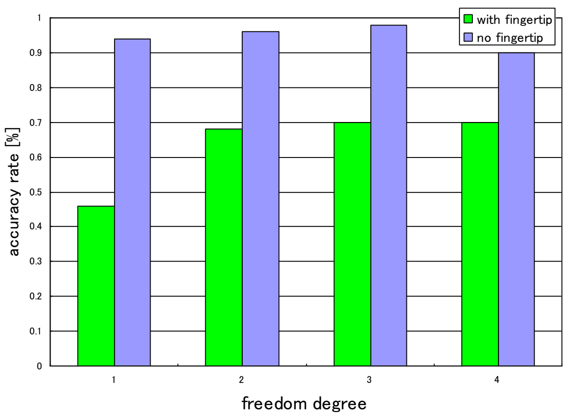


Figure 10. degree of freedom and accuracy rate

V. Grasping Motion And Human Brain Activity

We are trying to process experimental measuring and discuss human brain activity on grasping movement and cognitive task. So, we measured brain activity from the viewpoint of blood flow changes when subject performed grasping movement including reaching. Six subjects were healthy males who were right handed. They were asked to read and sign an informed consent regarding the experiment. In measurement, f-NIRS(Functional Near Infrared Spectroscopy) made by SHIMADZE Co. Ltd. were used.

A. Experimental Method

Subjects were asked to grasp the piece of wood, pointer, column-shaped metallic bar and hammer based on instructions from operator (Fig. 11). Brain activity was measured under four conditions. Subjects grasped objects actively with their eyes open (1) or close (2), and passively with their eyes open (3) or close (4). In addition, subjects were told to perform simple grasping motion or do it with imaging motion for using object.

Subjects took a rest during 10 seconds at least with their eye close before starting task and the time design was rest (5 seconds) – task (10 seconds) – rest (5 seconds). Finally, subject closed their eyes for 10 seconds again after task. Then, the brain activity was recorded from the first eyes-closed rest to the last eyes. The part of measurement was the frontal lobe.

B. Experimental Results

Fig. 12 shows one subject's measuring result. At the first, Hb-oxy was increased in overall frontal lobe after start of grasping task. This tendency was common among subjects. After that, Hb-oxy was increased and decreased in synchronization with task and rest. Also, there was remarkble tendency like this during task with imageing and their eyes open.

Analysis was performed one-sample t-test and sample was variation in brain activity during about four seconds after starting tasks.

As a result, there was not significant differences at frontal lobe. However, it was shown as common tendency among subjects that there was adifference in variation of oxy-Hb density due to presence or absence of existence or non-existence imaging motion and eyesight. It was thought that this result was derived from planning for grasping and visuals timulation.

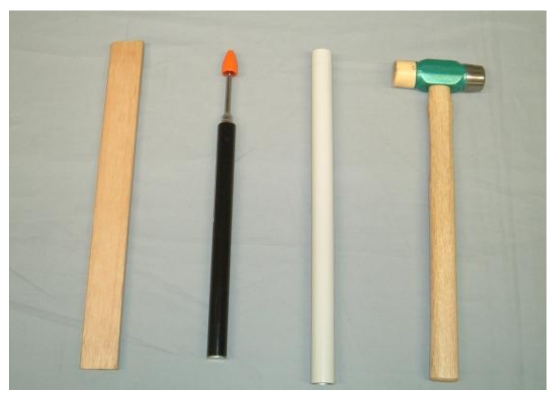


Figure 11. Grasping object

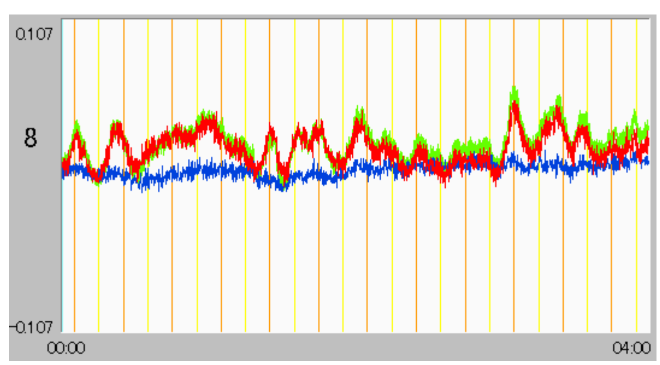


Figure 12. Measuring Result of grasping steering wheel

VI. Conclusion and Future Work

In this study, we performed three experiments. The analysis of human grasping movement is important in developing methodologies for controlling robots or understanding human motion programs. In analyzing human grasping movement, it is advantageous to classify movements.

This paper first studies that the grasping angle changes. From the results, it was determined that basically the wrist angle and finger baseline angle are proportionally related. In this report we attempted to classify the grasping patterns based on “purpose of grasping movement”. In the results described in Section III.B, we found two movements related to the “purpose of grasping movement”. Some grasping movements change when you give them a purpose before grasping, but some grasping movements do not change until the object is lifted and the action accomplished. These results suggest the possibility of the classification of grasping movements according to “purpose of grasping movement”. Furthermore, it would be useful to classify the grasping movements that have similar grasping forms or distributions, but have different actions after grasping. At a later stage, we would like to classify grasping movements according to the "purpose of grasping movement" and use the results of the distribution in grasping patterns to create more useful classifications for grasping movements in engineering.

The results in Section IV.B are important, in using the classification. The relation between accuracy rates basically increases with the degree of freedom, but if the degree of freedom increases with no useful feedback, the results would differ from expectation. Therefore, the challenge which lies ahead is to find effective ways to use the tactile information.

In terms of measuring brain activity, we plan to examine change of brain activity due to shape of hand and object as well as a review of experimental design.

References

[1] G Schlesinger: “Der Mechanische Aufbauder kunstlischen Glieder [The mechanical structure of artificial limbs]”, In M.Borchardt et al. (Eds.),Ersatzklieder und Arbeishilefen fuk Kriegsbeschadigte und Unfallverlezte, pp. 21-600, Berline: Springer, 1919.

[2] CL.Mackenzie and T.Iberall:“The Grasping hand,”G.e.Stelmach, P.A.Vroon(Eds.), Advances in psychology, no.104, pp. 15-46, Amsterdam, 1994.

[3] M.A. Arbib, T.Iberall, and D.M.Lyons: "Coordinatecontrol programs for movements of the hand," In A.W.Goodwin & I.Darian-Smith (Eds.), Hand function and the neocotex, pp. 111-129

[4] J.R.Napier: ”The prehensile movement of the human hand,” The journal of bone and joint surgery, vol.38 B, no.4, pp. 902-913, 1956.

[5] M.R. Cutkosky: “On grasping choice, grasping models, and the design of hands for manufactureing task, ”IEE Trans. Robotics and Automation, vol.5, no.3 , pp. 269-279,1819.

[6] N.Kamakura, M.Matsuo, H.Ishii, F.Mitsuboshi, and Y.Miura: “Patten of static prehension in normal hands,” The American journal of occupational therapy, vol.34, no. 7, pp. 437-445, 1980

[7] S.B.Kang and K.Ikeuchi: “Towed automatic robot instruction from perception – recognizing a grasping from observation,” IEEE Trans. Robotics and Automation, vol.9, No.4, pp. 432-443, 1993.

[8] M.Fujiwa, Y.Yokokohji, and T.Yoshikawa: “Dynamic Multi-Fingered Manipulability for Articulated Hands and its Extension to a Design Index for Master-Slave Hands,” Journal of the Robotics Society of Japan, vol.22, No.2, pp. 264-272, 2004

[9] K.Bernardin, K.Ogawara, K.Ikeuchi, and R.Dillmann; “A Sensor Fusion Approach for Recognizing Continuous Human Grasping Sequences Using Hidden Markov Models” IEEE Trans. Robotics, 21(1):pp. 47-57,2005.

[10] S.Shimizu, M.Shimojo, S.Sto, Y.Seki, A.Takahashi, Y.Inukai, and M.Yoshioka; “The Relationship between Hunam Griip Types ans Force Distribution Pattern in Grasping,” 8th International onferrence on Advanced Robotics, pp. 299-304, 1997

[11] K.Shiraishi, T.Kondo, and K.Ito; “Analysis on the Task Dependency of Human Grasp Movements and Application to Robot Control” SICE Symposium on Systems and Information,pp. 283-286,2002

[12] S. Matsuzawa, S. Shimizu, S. kikuchi, E. Watanabe, H. Kina, H. Hurata, N. Takahashi, and T. Ehara,“Proposal for a New Classification Method for Grasping Movement,” The 13th International Conference on Advanced Robotics, 2007, Jeju Island, Kores.

[13] N. Takahashi, S. Shimizu, Y. Hirata, H. Nara, H. Inoue, N. Hirai, S. Kikuchi, E. Watanabe, and S. Kato,“Basic study of Analysis of Human Brain Activities during Car Driving,” the 14th International Conferrence on Human-Computer Interaction, 2011, Orlando, Florida, USA.

[14] S. Matsuzawa, S. Shimizu, S. kikuchi, E. Watanabe, H. Kina, H. Hurata, N. Takahashi, and T. Ehara,“A Basic Study of Grasping Motion for Advanced Interface,” the 12th International Conferrence on Human-Computer Interaction, 2007, Beijing, China.

Analyzing the Effects of Virtualizing and Augmenting Trading Card Game based on the Player's Personality

Mizuki Sakamoto, Todorka Alexandrova, Tatsuo Nakajima
Department of Computer Science and Enginnering
Waseda University, JAPAN
{mizuki, toty, tatsuo}@dcl.cs.waseda.ac.jp

***Abstract*—In this paper, we focus on the Trading Card Game (TCG), which offers two versions of playing. One is the version of the TCG played with paper-based cards, and the other is the TCG played on a computer. We discuss the lost reality and the lost enjoyment by playing the computer based TCG compared to the paper based one. For analyzing the virtuality in the computer game, we propose a scenarios-based analysis based on the player's personality. We believe that the personality analysis described in the paper is useful to analyze the human social relationships in various games and social media, and also clarifies the possible obstacles and reasons for dissatisfaction in using social media and playing computer based games for each type of personality.**

Our study also claims that computer technologies can solve the problems caused due to the virtuality introduced in a game by using augmented reality techniques. We present *Augmented Trading Card Game* (*Augmented TCG*), and describe the results of some experiments with it that show how some problems and pitfalls of the remote trading card game play on the Internet could be improved. We believe that the case study and the scenario analysis given in this paper would be useful to improve the current realizations of games.

Keywords - Virtualization and augmentation; Personality; Game design; Augmented reality; Scenarios-based analysis.

I. INTRODUCTION

Our daily life becomes more and more virtual by making the surrounding daily artifacts more intelligent [6, 14]. Our definition of virtual is something that does not really exist, but has a real effect on our daily life as if it exists. *Jean Baudrillard* explains our consumption behavior as we consume symbols associated with things, not the things themselves [3]. Since the symbolization of things will accelerate by embedding computers in our life, our virtual consumption will progress rapidly.

The impact of introducing virtuality in our daily life has not been discussed well, although digital books have become very common everywhere, and virtual worlds like The Elder Scrolls V: Skyrim [26] are creating almost indistinguishable real worlds within a computer. Specifically, it is an important issue to discuss the effects of replacing real things with virtual things by using computers. In recent games, physical tangible things are replaced by virtual things in the computers, and multiple players can enjoy easily a remote game play without meeting each other. Even though the rules of a game do not change, the play style significantly changes according to the game's realization method. However, currently, there is no good way to analyze the effects of the different realization methods in a systematic way.

In this paper, we mainly consider the effects of virtualization in the *Yu-Gi-Oh!* trading card game [27] as a case study of the virtualization of daily objects. The *Yu-Gi-Oh!* trading card game (TCG) as shown in Fig. 1 is a turn based game that uses paper cards and is played in a one-to-one or two-to-two manner. A player collects cards by buying packs of cards that contain sets of randomly chosen cards or by exchanging cards with other players. *Yu-Gi-Oh!* cards are categorized into three types: monster cards, spell cards and trap cards. A player chooses his/her favorite cards, and constructs his/her own original deck, which consists of at least 40 cards. We call the battle with *Yu-Gi-Oh!* cards a duel. Each player starts a game with a certain number of points called life points and performs the duel by summoning his/her monsters, battling against the opponent player with his/her monsters or using spells and traps. Depending on the action taken and the outcome of it the life points decrease or increase. If the life points of a player become zero or the player cannot draw cards from his/her own deck any more, then that player loses the duel and the game ends.

Figure 1. Playing Trading Card Game.

There are several versions of the *Yu-Gi-Oh!* TCG that can be played on personal computers and allow players to play

 ISBN: 978-1-61208-250-9

the game remotely. The cards are represented in the computer virtually, and players manipulate the cards using a mouse. A single player can also play the game, and in such case the duel is performed “against the computer”, i.e., the game AI. In this paper, we focus on a version of the game that is played on *Nintendo DS* as a computer-based TCG. We call this version the *DS version* and it can also be either played by one person with the game AI as an opponent or by two people, who play against each other by connecting two *Nintendo DSs* through WiFi connection. Also, we call the version of TCG that uses paper-based cards *the paper version*.

For analyzing the player’s behavior in the *DS* and the *paper version*, we classify the *Yu-Gi-Oh!* TCG players based on the three personalities proposed by *Karen Horney* [7]. For each personality we first create a scenario that describes the typical behavior of the corresponding player. Then, we discuss how the scenarios change in case of the paper and the *DS version*. The *Yu-Gi-Oh!* TCG involves various sources of pleasure besides just playing, such as, completing collections of cards, structuring decks, communicating with other players, trading, battling and making links to *Yu-Gi-Oh!* TV animations and *Yu-Gi-Oh!* comics. However, players with different personalities focus on and enjoy different aspects of the game. That is why it is important to analyze the effects of the different realizations of the game for the different players’ personalities.

There are two main contributions of our research. The first contribution is to propose a scenario-based analysis of a computer game based on the player’s personality. In our research, we also show that *Yu-Gi-Oh!* TCG offers multiple ways to play the game, and these ways are identified by analyzing the player’s personality. We believe that the personality analysis described in the paper is useful to analyze the human social relationships in various games and social media, and also clarifies the possible obstacles and reasons for dissatisfaction in using social media and playing games for each type of personality. As a second contribution, our study claims that computer technologies can solve the problems caused due to the virtuality introduced in a game by using augmented reality techniques. We present *Augmented Trading Card Game* (*Augmented TCG*), and describe the results of some experiments with *it* that show how the pitfalls in the remote trading card game on the *DS version* could be improved. We believe that the case study and the scenarios analysis given in this paper would be useful to improve the current realizations of games.

The remaining sections are structured as follows. In Section II, we present some related work to our research. Section III introduces scenarios of trading card game plays based on three personalities, and in Section IV, we analyze the scenarios and how the virtuality is introduced when implementing the remote trading card game on the DS version. Section V presents *Augmented TCG* that augments the trading card game by using augmented reality technologies, and shows how the augmentation influences the game play according to the player’s personality. Finally, in Section VI, we conclude the paper and show some future directions.

II. RELATED WORK

In this paper, we adopt the *Karen Horney’s personality theory,* but we also consider using other personality theories to analyze the game. *Richard Bartle* analyzes the personality of a player in a role playing game [2]. He describes eight personalities: *opportunists, planners, politician, scientists, hackers, networkers, friends, griefers* as roles used in the game, and he claims that a player changes his/her roles according to the current situation. However, the classification does not take into account the human relationship that is important when analyzing trading card games. On the other hand, *Karen Horney’s personality* theory is based on the essential desires in the human relationship. In TCG, the human relationship is the most important factor to define the game. Thus, we consider *Karen Horney*’s classification as more appropriate to analyze TCG.

The *big five theory* claims that human personality can be analyzed by five factors: *openness, conscientiousness, extraversion, agreeableness,* and *neuroticism* [15]. The factors are demonstrated by rigorous scientific experiment. On the other hand, the Karen Horney’s personality theory is classified based on the human relationships like isolation, cooperation and conflict. At the moment, there is no psychological discussion to show the relationship between these two theories.

In [4], the scenarios analysis focuses on only one personality: moving toward people because the personality is the majority in our real world. We focus on more comprehensive analysis based on three personalities. One of the essential contributions of our paper is that players of trading card games are classified into three personalities, and the players of all personalities can enjoy the game.

The *product attachment theory* [16] explains that a user feels more empathy when the personality of a product matches the user’s personality. Applying the theory to TCG, TCG has three personalities and they correspond to the player’s personalities. Daily products also have multiple personalities [20], and our approach to analyze scenarios based on multiple personalities will be useful to design daily products.

There are several ways to augment the *Yu-Gi-Oh!* TCG. *Duel Accelerator* [24] is an online-based *Yu-Gi-Oh!* TCG where each player chooses an avatar that identifies him/her from the other players, and special effects on cards make players more excited to play the game. Also, the *Skype duel* uses Skype to show remotely each player’s card on the opposite player’s display and the voice communication between the two players is possible as well.

CyberOne [23] is a new TCG that enhances previous TCG. Each paper card has a sequence number. When the number is input, the corresponding virtual card appears in the online

 ISBN: 978-1-61208-250-9

TCG. The player can exploit the tangibility of cards, but he/she can also enjoy additional special effects, which is an example of the advantages of the virtual cards. Also, since the rules are simplified, the duel is automatically progressed without the player doing anything. However, the enjoyment of constructing an original deck carefully, remains for the player. Moreover, the two players do not need to play the game at the same time because they just need to construct their decks. One significant advantage of this type of TCG is that for a player with the moving away from people personality there is no obstacle to play a duel. However, since the TCG does not encourage the communication in the game, it may not be suitable for a player with moving toward people personality.

Augmented reality techniques may be used to enhance the existing games. For example, in [21], there are several augmented reality games that enhance traditional games. Specifically, *Augmented Go* [9] is a promising approach to maintain the advantages of the physicality, but to add virtuality to the Go board game. In [19], a pervasive game is designed with virtual and tangible objects, and two approaches are compared to investigate the effect on social interaction and physical activities. Also, in [5], Pacman and Ghosts are human players in the real world experiencing computer graphics fantasy-reality by using wearable computers on them.

The most significant difference between our approach and previous approaches is that we propose the usage of virtual characters from animation and game stories [18]. Specifically, this approach is very different from using avatars that identify users. The virtual characters in our approach add the fictional atmosphere of the stories to the trading card games and enhance players' experiences if players know the stories well.

The interaction between the virtual environment and the physical environment is one of the hottest topics in the game research. In [13], authors present an introduction and overview of the field of pervasive games that can blend real and virtual game elements. Our analysis can be extended to discuss how the games in virtual environments lose the reality, and to consider how the techniques of pervasive games can be used to improve the lost reality. Also, in [18], we discussed how a virtual character is effective when the virtual and real world are blended. Specifically, a virtual character used in animation and game stories can give us an enhanced feeling to seamlessly blend our real life and the fictional world of the animation and game stories.

III. SCENARIO ANALYSIS BASED ON PERSONALITY THEORY

The personality classification used in this paper is based on the *Karen Horney's personality theory* [7]. TCG offers several ways to enjoy the game, and giving a good classification of the players' personalities would allow us to analyze the ways to play the game in a more systematic way. We believe that the way and attitude of playing TCG for a certain player is identical to creating his/her own original and personal story. Even though each player emphases on a different way to play the game, it is possible to classify the players into few patterns. Therefore, creating scenarios based on the personality theory concepts is an appropriate approach to analyze the behavior and intention in the game play.

A. Player's Personality Analysis for Trading Card Game

Personality refers to how people evaluate the personalities of other people, including: 1: the characteristics or qualities that form an individual's character, 2: qualities that make someone interesting or popular (Oxford Dictionary, 2006). *Karen Horney* classifies human personalities into three types: *moving away from people*, *moving toward people*, *moving against people*. In this paper, we consider three scenarios based on this personality classification for *Yu-Gi-Oh!* trading card game players.

1) Moving Away from People: A person belonging to this category likes to seek completeness and perfectionism in a closed world, and attempts to establish his/her identity by self-satisfaction. Therefore, the person usually collects cards alone, and is really determined and concentrated to collect all the cards he/she wants. A player with this personality has a strong attachment to his/her cards, and the story to encounter the cards is very important for him/her. This means that he/she establishes his/her identity and deepens the attachment to the cards by knowing and exploring carefully the background of each card.

2) Moving Toward People: A person belonging to this category attempts to establish his/her identity by acquiring the approval and affection from others and seeking a partner. He/she likes the interaction with others by using cards, and constructs his/her original story by deepening the friendship with his/her friends and achieving self-realization together with others.

3) *Moving Against People:* A person belonging to this category attempts to establish his/her identity by obtaining social evaluation and praise from others, and creates his/her story by seeking his/her perfection. Also, he/she likes to be superior to the other party and satisfies his/her pride by showing his/her advantages to others.

B. A Scenario for Moving Away from People

"*Mizuki is a university student who likes Yu-Gi-Oh! cards. Her favorite deck type is Light Sworn. The illustrations on the cards of the Light Sworn series are very lovely and pretty to her, and she really enjoys just arranging her cards and looking at them. However, she still does not have all the cards from the Light Sworn series. Specifically, she does not have the Lightsworn Sorceress Lyla card, but she really wants to have it. Lyla is a beautiful lady, and the card*

 ISBN: 978-1-61208-250-9

depicting it is twinkling as well, which makes it even more attractive. Moreover, the card of Lyla would make Mizuki`s deck stronger. But, the card is extremely rare, and it is very hard to get it. She has the chance to get the card as a used one, but she does not like the idea since she would not feel the card as hers. However she is determined to find the card and for that purpose she goes to the card's shop and buys two boxes of cards that may contain the Lyla card. Each box costs 4500 yen and contains 30 packs, where each pack contains 5 cards.

After coming back home, Mizuki impatiently starts opening the packs. She is so excited, her heart beats fast, and she carefully opens the packs, one by one. She hopes to get the card whenever she finds a card whose edge has the same color as the Lyla card. Unfortunately, just the edge of the card is the same color and the illustration on the card is a different monster. She is very disappointed not to get the desired card again and again. Unfortunately, she could not find the Lyla card in the two boxes she bought this time. Although she is disappointed, she carefully sorts and keeps all the cards bought today in her collection. She could get some other rare cards this time as well, but she could not be very happy about that. She really wants to get the Lyla card, so she decides to buy seven more boxes and look for it. No matter how expensive it is, she will continue buying boxes until she gets the Lyla card.

Again, in the first box, she could not find the Lyla card. She also could not find the card in the second box. Now she thinks she will not be able to find the card this time as well and feels desperate. But, finally, she could find the Lyla card in the third box. Mizuki becomes really happy, inserts the Lyla card very carefully in a sleeve and adds it happily to her deck. However, she has no plan to use the deck to play Yu-Gi-Oh! TCG, but just to keep it for herself. She continues to open the fourth and the fifth box still excitedly, and finds another Lyla card in the fifth box. Thus, she becomes super happy again. After opening all the boxes, she is able to complete fully her collection of cards in the Light Sworn series and is very satisfied about that. Although she gets more than ten same cards, she is really happy to have completed her goal. ".

C. A Scenario for Moving Toward People

"*Toty has started to play the Yu-Gi-Oh! TCG with the recommendation from her friend Eiji. Since Eiji has more experience with the Yu-Gi-Oh! TCG than her, he teaches her how to play and enjoy the game. Eiji takes her to a card shop near their homes. Eiji plays the game with the deck of the Elemental Hero series, and he wants to get cards from the Elemental Hero series. Each of them buys 10 packs. Then, they go back to Toty's home together and start opening their packs. There is a super rare card in Toty's Elemental hero series pack that he really wants to have. Eiji negotiates with Toty to exchange the card with him for another super rare cards that he currently has and she agrees since she does not consider to use the card in her deck. Toty passes the card, and asks Eiji to exchange cards with her again when he gets a card that she wants to have in the future. Then, finally, they start to play a duel.*

Although Toty really loves playing the duel, she does not really understand the game's rules yet. Thankfully to Eiji's kind help and explanations after playing several duels, Toty increases her playing skills and understands how to use the super rare card she got from Eiji. They play ten duels, in which Toty has three wins and seven losses, but in all the three wins, the super rare cards that she has got from Eiji plays a key role. Now, it becomes dinnertime, Eiji returns to his house and both of them are looking forward to their next duel.

At home, Eiji starts to rearrange his deck by selecting some more suitable cards for dealing with the super rare card that Toty has in the next duel. He raises his fighting spirit and says to himself "Well, that deck will not lose in the next duels with Toty.

After dinner, Toty also looks carefully at the cards she has got from Eiji. She has enjoyed today's duels and chats with Eiji a lot and is satisfied with the score, but she promises to herself to win the duels with Eiji next time.

They both had a great day and are really excited and looking forward to their next duels and talks about Yu-Gi-Oh! Cards".

D. A Scenario for Moving Against People

"*Tatsuo is more than 10 years advanced Yu-Gi-Oh! TCG player. For him, wining the duel is the most important issue, and he is very interested in increasing the power of his deck. Tatsuo has also participated in tournaments a couple of times. Today, he visits again one of his favorite cards shop and plays duels there. Since he visits the cards shop every week, he has some friends there to play with. They all are advanced players, and use a variety of decks and tactics. In their play, all decisions are completely effective. All of them know the effects of the cards by just looking at the illustrations of the cards, without reading any of the explanations. They all play the duels very seriously and concentrated without even chatting during the play. Specialized words like: "Draw", "Summons", "Attack", "Trigger Monster Effects", "Trigger Magic Effects", "Trigger Trap Effects", "Chain", "Turn End" are the only communication among them during the play. Although Tatsuo is a strong player, it is not easy for him to win the duel today. Just one mistake would make him lose, so he feels great thrill for the play. He needs to predict the opponent player's tactics from the opponent player's field and graveyard and from the opponent's gaze and face expression, and choose the most effective card towards it. Finally, he wins the duel and feels extremely happy and satisfied. Then, his friend starts a new duel with another player after changing his deck with a deck, which would be more effective for the new opponent player's deck.*

Some of his friends start to return to their homes and Tatsuo also decides to quit playing for today. He has lost two duels today. He analyzes the reasons to lose, and considers how to improve his play next time by choosing more appropriate and stronger card. Immediately, he tries to buy that card as a used card before coming back home. Only in case he cannot find the card as a used card, he will try to buy some new packs and look for that exact card in them. In such case, if he cannot find the target card in the packs, he will sell all the cards at the shop. Tatsuo will definitely come to the shop next week, and try to win all duels."

IV. THE INFLUENCE OF VIRTUALITY THROUGH SCENARIO ANALYSIS: LOST REALITY

In the previous section, we have presented three scenarios for the case when players use real paper cards in the game. In this section we consider how scenarios change in the case of the *DS version* of the TCG that uses digitally implemented virtual cards. In this paper, we call *lost reality* the situation in which a player loses some of the realities, he feels during the play with the real paper cards, while playing the *DS version*. In this section, we consider the lost reality issues in the scenario of each personality.

A. Moving Away From People

In the scenario described in Section III.B, the physical and visual value of the real paper card is essential for a player with this personality. As described in the scenario, the player can achieve a feeling of fulfillment just by the arrangement of the cards and by looking at them. Also, she feels that the twinkling super rare cards are very precious for her. The digitally implemented virtual cards are difficult to offer detailed visual information and a sense of twinkling. Also, in the scenario, the player obtains more than 10 same cards. Virtual cards lose the sense of quantity because the quantity is represented just as a digital number, while real paper cards offer the sense of the quantity as the thickness of the set of cards. Buying boxes of cards, opening packs and keeping and arranging cards in sleeves carefully is related to the sense of touch. However, even though virtual cards lose all these physical senses they have some advantages as well. Virtuality makes it easy to manage the cards, i.e., it is easy to sort the cards and keep them together. But, virtual cards may make the collection of cards a boring task.

Virtual economy is also one of the ways to recover the reality [11]. In the *DS version*, *duel points* (*DP*) are added to the game. A player can get cards by using his/her *DP*. *DP* can be increased in various ways in the game such as winning a duel with the game AI. Thus, without spending money on collecting cards, a player can obtain the cards he/she likes using his/her *DP*. Therefore, the will to increase his/her *DP* winning percentage encourages a player to play the game with the game AI or other player through a WiFi connection. A player with the moving away from people personality is discouraged to play the game when the card is virtual because virtual cards decrease the player's motivation to collect the cards. Introducing virtual economy increases the values of the cards, which offers the possibility to recover the motivation to play the game.

B. Moving Toward People

In the scenario described in Section III.C, the player with the moving toward people personality considers that the relationship and the communication with other people are the most important conditions to enjoy the game. Going to a nearby cards shop together with a friend, visiting a friend's house to play together and opening packs of cards together are some typical important actions in the scenario. It is not easy to buy cards and open packs together when using virtual cards. Especially, in the *DS version*, each player does these actions alone. Also, changing the rules, such as exchanging a card for two, is a limited action since it is hard to change the rules in a flexible way in the *DS version*. Moreover, it is impossible to chat with friends when playing the *DS version* remotely against them. Also, while playing a duel, it is difficult for friends to support each other and give advices to each other.

As described above, a play with real cards involves a duel that is performed by players located at the same place, who can communicate with each other directly. On the other hand, a player performing the game with virtual cards may lose the chance to enjoy the communication with the opponent player. However, if the two players are located at the same place, then even using the DS version of the game will not bring any communication difficulties to them, since they can see and talk to each other. In [12], there is a similar discussion for the case when online RPG game is played among people who are located at the same place. Of course, the sense of physical presence increases the pleasure of the game, but it is more important whether the two players can use a rich communication channel or not.

A player with the personality of moving toward people would be mostly motivated and excited to play a duel with his/her opponent face-to-face.

C. Moving Against People

In the scenario described in Section III.D, for the player with the moving against people personality, wining the duels is the most important pleasure and goal in the game. Since winning against real human is different from winning against game AI, a player of this type is much more satisfied to win duels against strong human players. However, one worth mentioning feature of computer games is the fact that it is possible for a player to reset the game by simply turning off the computer if the current situation and score are not convenient and favorable for him/her. In the scenario, the player feels a strong sense of thrill because just one mistake in his play would make him lose the duel, while in the virtual world, it is easy to reset the game when the game does not advance according to his preferences. Thus, in the

 ISBN: 978-1-61208-250-9

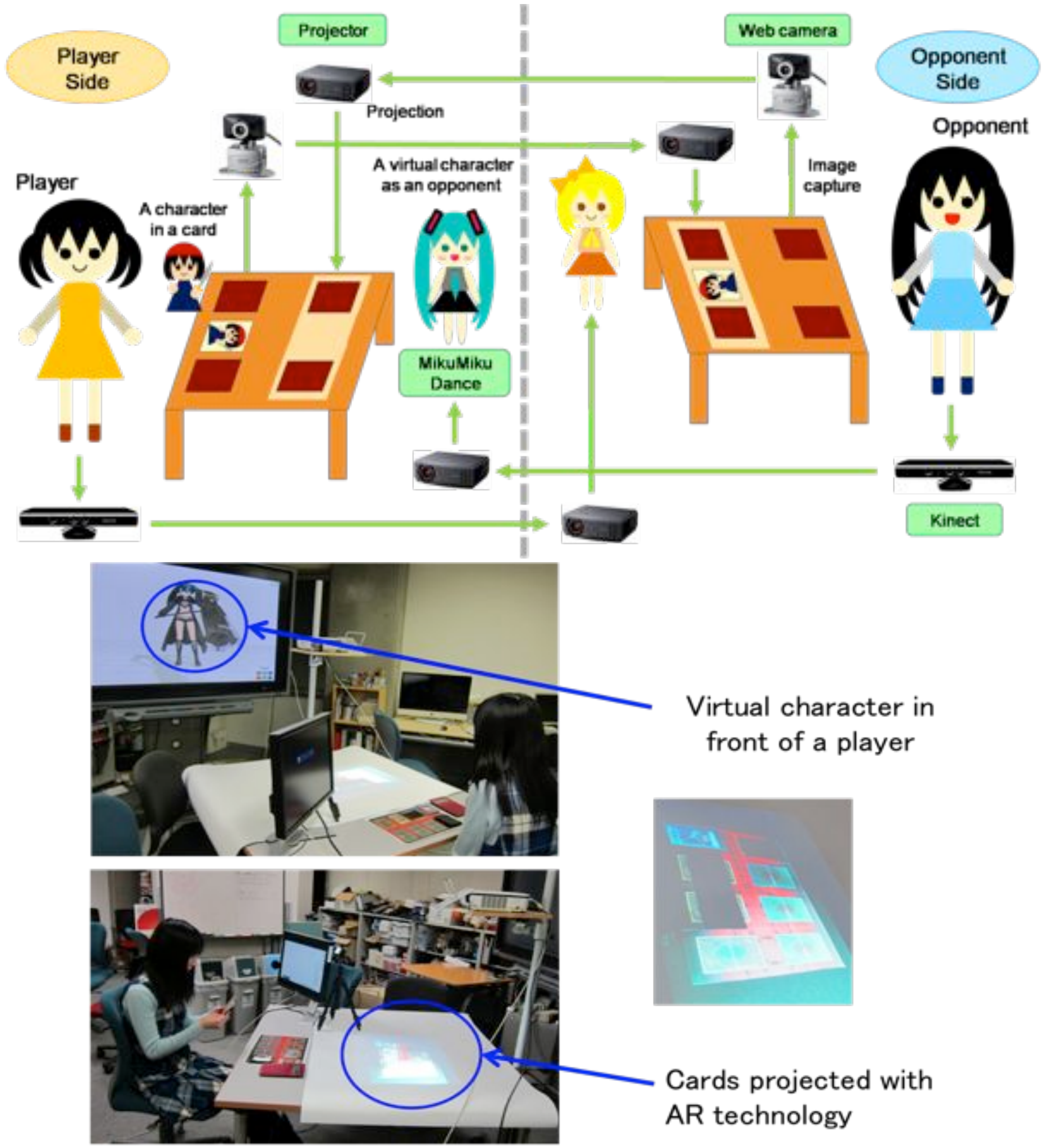


Figure 2. Augmented Trading Card Game.

virtual world, the sense of thrill, challenge and risk is decreased with the possibility to terminate or reset the game at any time.

Also, in the scenario, the player predicts the opponent player's tactics and move by observing his/her gaze and face expressions. This reality is completely lost when players cannot see each other, which decreases the excitement and enjoyment of the game, as well as the satisfaction of winning the game. Moreover, since the player does not see the opponent player, he/she tends to easily cheat in the game. For example, in the *DS version*, the winning percentage is recorded, and this becomes a big incentive for the players with this personality to win the duels. Being determined to win the game, many players using the Nintendo DS, tend to disconnect the connection with the remote opponent player once they realize they might be losing the game. Currently, this becomes a serious problem for players who play the *DS version*, which shows that it might be difficult to maintain a good moral ethics in case the sense of presence of the opponent player is lost.

V. AUGMENTED TRADING CARD GAME

A. Motivation and Overview of Augmented TCG

As described in the previous section, if the reality of a game is lost, the incentive to play the game and the enjoyment of the play decrease. For a player with the moving away from people personality, the lost reality of the physical sense of the cards decreases his/her motivation to collect cards. Also, for a player with the moving toward people personality the incentive to play a game when the communication between players becomes difficult or inactive, diminishes. And for a player with the personality for moving against people, the motivation to play a game weakens if it is impossible for him/her to look at the face of the opponent player in order to observe his feelings and thus predict his tactics and moves. Moreover, such player does not feel the self-esteem when other people do not see his/her victory. This means that it is hard for a player to create a story based on his/her own memories by the lost reality.

When analyzing a player's intention and behavior, behavior economy plays an important role. Behavior economy demonstrates that human intention and behavior are greatly influenced by a frame defined for each condition [1]. This means that a frame defines how to offer pleasure to a player and also changes according to the realization of the game. In this section, we consider a new frame to support remote TCG play that covers all the personalities. The frame also offers the physical tangibility of paper cards. The case study will show our personality analysis is effective to analyze the pitfalls of the current game design.

Jean Baudrillard proposed a concept called hyper reality [3]. Augmented reality technologies are used to introduce virtuality in a game, but also they are used to increase the hyper reality that is lost by the virtuality. In this section, we present *Augmented TCG*, which offers one possible frame to play the *Yu-Gi-Oh!* TCG that recovers the realities lost in the *DS version*.

We have designed *Augmented TCG* taking into account our pre-observations and analysis of the scenarios described in the previous section. The basic design approach is similar to the one in the augmented reality games introduced in [21]. As shown in Fig. 2, the two players are located in different places. Each player's cards (real paper cards), in his/her duel field on the table in front of him/her, are captured by a camera and projected on the opponent player's table. Also, each player is represented as the 3D model of a virtual character used in popular animations and games, and this character is shown to the opponent player. In the current implementation, *MikuMikuDance* [25] is used to show the 3D models of the virtual characters. *MikuMikuDance* is free software for creating 3D movies by using virtual characters. The virtual character is controlled using *MS Kinect* and its movement is synchronized with the movement and the behavior of the player it represents. Moreover, the two players can communicate with each other via Skype if desired, and thus it is possible for them to introduce each other directly instead of using virtual characters. This option will allow us to compare the case of the players communicating with each other through virtual characters and the case of the players directly communicating with each other.

Fig. 2 also shows a basic configuration of *Augmented TCG*. The most important design issue of *Augmented TCG* is the approach to use real paper cards not virtual cards, implemented digitally in a computer as shown in the figure, and the player performs the duel with a virtual character as an opponent. The system is useful to discuss the issues how virtual and real components in the TCG game affect the player's game play.

B. Experiments and Design Implications

We have conducted a user study to show some experiences and observations with *Augmented TCG*. We have recruited five participants who are familiar with *Yu-Gi-Oh!* TCG, have more than three years experience with *Yu-Gi-Oh!* TCG, and know both the *paper version* and the *DS version* well. Three kinds of experiments have been performed for each participant and each participant performed the duels in the experiments against one of the authors of the paper, who has deep knowledge about the TCG and could lead and control the experiment so that all participants play the game under the same conditions. After the experiments we made an interview with the participants through semi structured focused group. In this subsection, we explain an overview of each experiment, and show some experiences extracted during the user studies.

In the first experiment, each participant played a duel against an opponent, represented by a virtual character, whose movement was synchronized with the movement of the opponent player by using *MS Kinect.* After the experiment a participant told us that if the virtual character was his favorite character, he could enjoy the duel better. Another participant said that he focused mostly on the opponent player's card, and presenting the virtual character was not important for him. However, most of the participants claimed that the virtual character was not enough to feel the presence of the opponent player because *Kinect* was unable to extract face expressions and eye movement. Also, the movement of the virtual character was not natural because *Kinect* interpreted the movement of the player based on only few body points.

We also compared the case of the player directly seeing his/her real opponent player with the case of using a virtual character to represent the opponent player as shown in Fig. 3. From the interview we have found that if a participant knows the opponent player well, then the real person view is more preferable and increases the reality of the game, while in the case of the opponent player being a stranger, some of the participants claim that using virtual character is preferable because they do not feel comfortable either showing themselves or seeing their opponents. A participant, whose personality is the moving away from people type, does not like to see the opponent player directly if he/she does not know the player well. For him/her, representing the opponent player as a virtual character is helpful to increase his/her motivation to play the duels. A participant with the moving toward people personality would prefer to see the real view of the opponent player. Therefore, the right representation of the opponent player depends on each player's personality. Finally, for a participant whose personality is moving against people the real gaze and face expression of the opponent player is essential in order to predict his/her tactics and consequent moves. That is why in such case it does not matter whether the opponent player is real or is represented with a virtual character, but the most important is that if he/she is represented by a virtual character then that character should offer a gaze and a face expression exactly the same as the real opponent player's one at that moment. The discussion teaches us that the reality of the virtual character is essential to satisfy and motivate all personalities.

Figure 3. Playing against a Real Opponent Player.

In the second experiment, the card that the opponent player draws is shown on a small display near the participant as shown in Fig. 4. As described in the scenarios presented in the previous section, virtual cards significantly decrease the motivation and the game enjoyment for players with the moving away from people personality. However, if the opponent player's cards are projected on the table, it might be hard to clearly see and understand the characters on them if they are too small or with low resolution and thus more difficult for a player to make a right decision. That is why in our settlement we show the card drawn by the opponent on a small display near a player. However, the necessity of such details strongly depends on the personality of a player. If a player's personality is moving away from people type, then he/she usually likes to win a duel because it is important for her/him to show the superiority of his/her favorite deck, but he/she may not have enough knowledge about other players' cards. Hence, he/she needs to see enough information of the opponent player's cards in order to play well. For a player with the moving toward people personality, winning a duel is not the most important point but the fact that he/she would be able to communicate and enjoy a game with a friend. In such case, the offered information by the displayed card is not so important since he/she can always ask his/her friend opponent for more information if necessary and such communication would even strengthen their friendship. A player with the moving against people personality has a lot of knowledge for most of the cards and just seeing the shape of the illustration on the card is enough for him/her to know its functionality. Thus, for him/her, it is not so important to show the small details of the cards. However, we believe that the possibility to show detailed card information would have a positive impact on the enjoyment of the game for each personality.

Figure 4. Showing a Card in a Small Display.

In the third experiment as shown in Fig. 5, while playing the game, another virtual character depicted on one of the player's cards appears on a small display near the player once that card is drawn out of the deck, and supports and encourages him/her to win the game until the end of the game. After the experiment one of the participants said that it was desirable that the card depicting the character shown on the small display did not lose from the attack of the opponent player. Another participant, who was not interested in the character shown, told us that it would have been more enjoyable if his/her favorite character encouraged him/her. Most participants said that the presence of the character was enjoyable, but it was hard to consider winning the game just from the character's encouragement. The participants' comments showed that they were aware that exactly the character depicted on one of their cards appeared on the small display without them being informed in advance about this feature of the system. A player with the moving away from people personality feels the importance to show the character depicted on his/her favorite card.

 ISBN: 978-1-61208-250-9

Especially, a female player likes a card depicting a lovely character and the approach is very effective for encouraging her to win the duel. For a player with the moving toward people personality, showing such a character is also a kind of special effect for encouraging him/her in the duel. So, if the character shows the current state of the duel, i.e. wining or losing, it may increase the pleasure of the duel. Therefore, the effect should be customized for each player's personality, but we believe that offering the encouraging effect to the duel would be useful for all personalities.

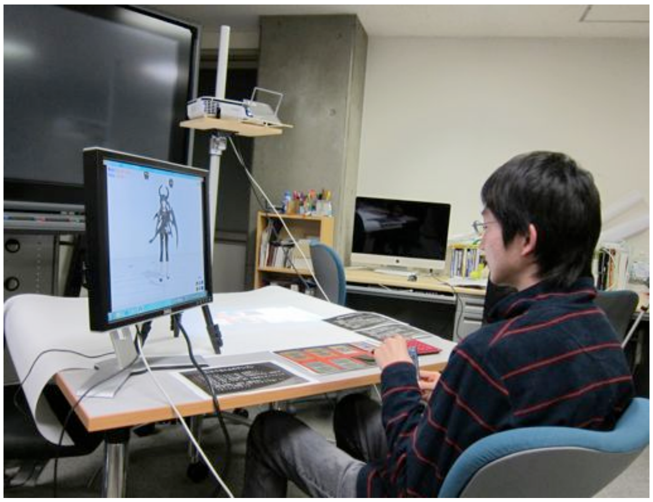


Figure 5. Encouraging a Player by a Virtual Character.

C. Discussions and Future Directions

In this subsection, we give some discussions on the observations and results obtained from the experiments with the *Augmented TCG.* In *Augmented TCG*, the physicality of the cards is the most important design decision, since virtual cards do not offer the sense of tangibility and thus lead to the loss of the reality in *Yu-Gi-Oh!* TCG. However, physical tangibility is a very important concept to offer enjoyment of the game to a player [8, 10]. Especially, for a player with moving away from people personality, collecting *Yu-Gi-Oh!* TCG cards is the most significant fount of pleasure, but virtual cards lose the sense of ownership and thus decrease the enjoyment and satisfaction of the game. However as a positive observation, adding virtuality may help to increase the empathy on the cards. For example, showing the special character depicted on one of the cards on a display near the player increases the empathy with that character. If the character were the player's favorite character then playing together and communicating with the character during the game would make the game even more enjoyable and exciting. Furthermore, even in case the card depicting that character is a very rare one, the player would make a great effort to get the card and enjoy the interactive play with the character.

For a player with moving toward people personality, feeling empathy with the virtual character representing the opponent player is important to maintain the incentive to play the game. In this case, the sense of the remote player is an important issue. If the player knows which virtual character his/her friend likes, using that character to represent the player increases the empathy during their play because he/she considers playing with his/her friend. The effect is similar to using an avatar. If a player knows the avatar of his/her friends well, he/she does not distinguish the avatar from his/her friends. However, using virtual characters from games and animations may enhance the conversation between a player and his/her friend if both are interested in the stories of these characters.

A player with moving against people personality feels more excited to play against *Yugi* or *Kaiba,* which are very skillful players in the *Yu-Gi-Oh!* animation story, because that player has a strong motivation to win the game against an expert player. Then, he/she feels the illusion of playing the game inside the animation story. However, in this case, as described in the previous section, the reality of the virtual character is essential. *Augmented TCG* synchronizes the movement of the remote player with the movement of the virtual character, but to maintain the reality of *Yugi* and *Kaiba,* the movement needs to be modified to reflect their natural characteristics in the animation story.

In our recent projects, analyzing the personality becomes a common tool when designing a new service. We especially believe that the analysis based on a user's personality can be used when designing social media. Recently, social media like *Twitter* and *Facebook* become very common as a tool to connect people and increase human creativity. However, from the aspect of the personality theory, the current social media does not support well people with the moving away from people personality. The scenarios analysis described in the paper is useful to improve the current social media for a larger number of people. Specifically, the personality analysis framework described in the paper would be useful to analyze the social aspect of any services, and we hope that the proposed approach can contribute to the progress of the future social media and services.

VI. CONCLUSION AND FUTURE DIRECTION

In this paper, we discussed the *Yu-Gi-Oh!* Trading Card Game, which offers two versions of playing. One is the version of the TCG played with paper-based cards, and the other is the TCG played on a computer. We discussed the lost reality and the lost enjoyment by playing the computer based TCG compared to the paper based TCG. For analyzing the effect of the virtuality, we proposed a scenarios-based analysis of a computer game based on the player's personality. Our study also presented that computer technologies can solve the problems caused due to the virtuality introduced in a game by using augmented reality techniques. We introduce *Augmented Trading Card Game*, and describe the results of some experiments with *Augmented TCG* that show how the pitfalls in playing trading card game among remote players on the Internet could be improved.

Computer games attract people by creating the illusion of being immersed in an imaginative virtual world with spectacle computer graphics and sound. The goals of computer games are more interactive than that of traditional games, and it brings players a stronger desire to play the game. Computer games designed with an optimal level of information complexity can provoke players' curiosity. Therefore, computer game intrinsically motivate players by bringing them more fantasy, challenge, and curiosity, which are the three main elements contributing the fun in games. Augmenting and virtualizing our daily life by using computer games is a promising way to extend traditional gamification. Our approach is also useful to analyze the lost reality of augmenting and virtualizing our daily life. We need to discuss these issues in our future research.

As a next step, it is necessary to identify metrics to analyze the scenarios more systematically. In [17], we identify five values to analyze the values of products and services and would like to investigate whether these metrics can be used to analyze the described scenarios. We need to consider how each value is related to the player's personality. We may change the values dynamically according to the current player's personality. Also, we are interested in designing gamification scenarios [22] based on the personalities. Each player's personality requires different incentives to motivate a user of the gamification service. That is why, future gamification frameworks need to take into account how to successfully incentivize players depending on their specific personality and we believe that the personality analysis given in this paper would be helpful in that direction.

REFERENCES

[1] Dan Ariely, "Predictably Irrational: The Hidden Forces That Shape Our Decisions", Harper, 2009.

[2] Richard Bartle, "*Virtual Worlds: Why People Play"*, MMP Game Development

[3] Jean Baudrillard, *The Consumer Society: Myths and Structures,* Sage Publications Ltd, 1998.

[4] Bandai Character Research Center, "*Card Game as a Drama to Create with Friends*, Dandai Character Research Center Jidai Report No.2, 2001, http://www.chara-labo.com/report2.pdf. (September 28 2012)

[5] Adrian David Cheok, Kok Hwee Goh, Wei Liu, Farzam Farbiz, Sze Lee Teo, Hui Siang Teo, Shang Ping Lee, Yu Li, Siew Wan Fong and Xubo Yang, "*Human Pacman: a mobile wide-area entertainment system based on physical, social, and ubiquitous computing*", In proceedings of Advances in Computer Entertainment Technology 2004, pp.360-161, 2004.

[6] Wimer Hazenberg and Menno Huisman, "*Meta Products: Buidling the Internet of Things"*, BIS Publisher, 2011.

[7] Karen Horney, *"Self-Analysis"*, W W Norton & Co Inc; Reissue, 1994.

[8] Hiroshi Ishii and Brygg Ullmer, "*Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms*", In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.234-241, 1997.

[9] Takahiro Iwata, Tetsuo Yamabe and Tatsuo Nakajima, *"Augmented Reality Go: Extending Traditional Game Play with Interactive Self-Learning Support"*, In Proceedings of the 17th IEEEE Conference on Embedded and Real-Tie Computing Systems ad Applications, pp.105-114, 2011.

[10] Patrick W. Jordan, "Designing Pleasurable Products: An Introduction to the New Human Factors", Routledge, 2002.

[11] Vili Lehdonvirta, *"Virtual Consumption"*. Publications of the Turku School of Economics, A-11:2009.

[12] Silvia Lindtner, Bonnie Nardi, Yang Wang, Scott Mainwaring, He Jing andWenjing Liang, *"A Hybrid Cultural Ecology: World of Warcraft in China"*, In Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, pp.371-382, 2008.

[13] Carsten Magerkurth, Adrian David Cheok, Regan L. Mandryk and Trond Nilsen, "*Pervasive Games: Bringing Computer Entertainment Back to the Real World*", ACM Computer in Entertainment, Vol.3 , No. 3, 2005.

[14] Tatsuo Nakajima and Vili Lehdonvirta, *"Designing Motivation in Persuasive Ambient Mirrors"*, Personal and Ubiquitous Computing, Vol. 17, No. 1, pp.107-126, 2013.

[15] Daniel Nettle, "Personality: What Makes You the Way You Are", Oxford Univ Pr., 2009.

[16] Mugge Ruth, Jan P. L. Schoormans, and Hendrik N. J. Schifferstein, *"Product Attachment: Design Strategies to Stimulate the Emotional Bonding with Products,"* in Product Experience, Elsevier, pp.425-440, 2007.

[17] Mizuki Sakamoto, Tatsuo Nakajima and Todorka Alexandrova, *"Digital-Physical Hybrid Design: Harmonizing the Real World and the Virtual World",* In Proceedings of the 7th International Conference and Workshop on the Design & Semantics of Form & Movement, pp.211-222, 2012.

[18] Mizuki Sakamoto, Todorka Alexandrova and Tatsuo Nakajima, "Augmenting Remote Trading Card Play with Virtual Characters used in Animation and Game Stories – Towards Persuasive and Ambient Transmedia Storytelling –", In Proceedings of the 6th International Conference on Advances in Computer-Human Interactions, 2013.

[19] Iris Soute, Maurtis Kaptein and Panos Markopoulos, "*Evaluating Outdoor Play for Children: Virtual vs. Tangible Game Objects in Pervasive Games*", In Proceedings of the 8th International Conference on Interaction Design and Children, pp.250-253, 2009.

[20] Jane Fulton Suri, *"Thoughtless Acts?"*, Chronicle Books, 2005.

[21] Tetsuo Yamabe and Tatsuo Nakaima, "*Playful Training with Augmented Reality Games: Case Studies towards Reality-Oriented System Design",* Multimedia Tools and Applications, Vol. 62, No. 1, pp.259-286, 2013.

[22] Gabe Zichermann and Christopher Cunningham, "Gamification by Design: *Implementing Game Mechanics in Web and Mobile Apps*", O'Reilly, 2011.

[23] CyberOne TCG, http://www.c-1cyberone.com/official_top/official_top.html (September 28 2012)

[24] Duel Accelerator, http://www.yugioh-online.net/v3/newvisitors/. (September 28 2012)

[25] MikuMikuDance, http://www.geocities.jp/higuchuu4. (September 28 2012)

[26] The Elder Scrolls V: Skyrim, http://www.elderscrolls.com . (September 28 2012)

[27] Yu-Gi-Oh! Trading Card Game, http://www.yugioh-card.com/en/. (September 28 2012)

Influence of Relationship between Game Player and Remote Player on Emotion

Masashi Okubo and Tsubasa Yamashita
Dept. of Information Systems Design
Doshisha University
1-3 Tatara Miyakodani, Kyotanabe, Kyoto, Japan
mokubo@mail.doshisha.ac.jp

Mamiko Sakata
Department of Culture and Information Science
Doshisha University
1-3 Tatara Miyakodani, Kyotanabe, Kyoto, Japan
msakata@mail.doshisha.ac.jp

Abstract— **The main trend in video game playing is the move to collaborative from competitive styles. Additionally, the game system is no longer confined to a stand-alone video game machine, but opened to other players through the Internet, so the competitors and the cooperators in the video game's world are not only programs, but also human beings. Over the internet, the players compete for a goal with their competitors in the competitive style game and obtain their goal together with their partners in the collaborative style game. Therefore, the game player cannot estimate the partner or enemy's behavior, and the interest in the game doesn't only depend upon the game itself but also on the partners' and competitors' behaviors. In this paper, we will investigate the influence of a partner or competitor on the performance and the state of mind of the game player. Another aspect of game playing is discussed by Mihaly Csikszentmihalyi, who identifies 'Flow', in which a person is fully immersed in a feeling of energized focus, as a central experience for enjoyable something. It is said that a person can feel Flow when they recognize their skill is just enough to accomplish the task. We will investigate the game player's state of mind based on Flow theory. As a result of some experiments, it will be shown that the player tends to feel good when the performance of their partner or enemy is almost as high as the skill of the subject. These experiments were performed by participants playing various video games, however, we think that the results of the experiment may be applied on the many and varied systems which support human motivation.**

Keywords; social game; video game; flow theory; collaboration; competition

I. INTRODUCTION

Recently, because of the diffusion of the Internet into the home and onto hand-held devices, Network Games such as Social Games have become popular. Further, it is not only competitive type games in which the player brings down an opponent, but also cooperative type games in which the player obtains the goal in cooperation with online partners that have grown in popularity. In this study, we will investigate the influence of the difference between competitive and cooperative type games as well as the opponent's effect on the primary player's state of mind and his performance.

The concept which is named Flow is widely acknowledged in Sociology. As mentioned previously, Flow is advocated by the social psychologist Csikszentmihalyi [1], and is the state in which a game player is fully immersed in a feeling of energized focus. A person can feel Flow in the condition, when he recognizes his skill level is just enough to accomplish the task, which may also be adjusted by changes in the challenge level as a player gains experience. We think that the video game player can feel Flow easily because the game player can raise his skill level by simply playing the game and the game system can raise the game level based on the player's ability automatically. In this paper, we also wish to discuss the possibility of the game system which induces Flow on the player's state of mind.

Until now, many varied game systems have been developed and have succeeded in the video game market. Some researchers in this field have focused on motivating workers by using video games, and enhancing the elements of entertainment and game design that have this effect [2]. Other research has investigated the influence of a player's skill and competitor's behavior on the player from the viewpoint of brain activity by using fNIRS [3]. Still, other researches on game systems based on Flow theory have been performed using commercial games [4]-[6]. There is a possibility that another results were obtained with another games. And this study will also examine game systems and human computer interaction based on Flow theory.

In this paper, the goal is to propose a game system which induces Flow and to experimentally measure this experience based on the influencing factors of *game type* (cooperative vs. competitive) and *opponent's ability* (lower/equal/higher). It is difficult to control these situations by using the commercial game. Therefore, It is necessary to design the experimental game system in which the game situation can be controlled by the experimenter. In Section II, Flow theory is introduced, and the relation between the competitive type or cooperative type games and that theory is discussed. Section III describes the experimental system which allows the experimenter to control the game situation and the remote player's skill, and experimental method to investigate the influence the remote player's skill on the game player's performance and state of mind. The experimental results are demonstrated in Section IV. Conclusions and future works are demonstrated in Section V.

Eventually, we think the knowledge gained from this research can be utilized in various kinds of computers and machine interfaces.

II. Flow Theory

A. Overview

Fig. 1 shows the model of flow state, which is advocated by Csikszentmihalyi. When the task challenge level, which means the difficulty of the task, is higher than the operator's skill, he probably feels 'Anxiety' and/or 'Stress'. Conversely, in the case that the challenge level is lower than the operator's skill, he will feel 'Relief' and/or 'Boredom'. In the case that the operator recognizes the challenge level is just right for his skill, he will feel 'Flow'. Additionally, when both levels are high and well balanced, then he feels it more strongly. It is said that the components of 'Flow', are as follows [1],[7]:

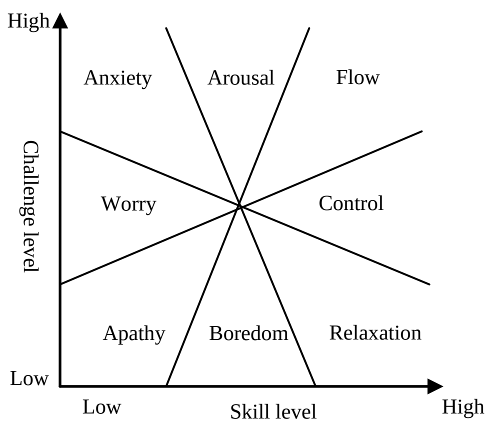


Figure 1. Model of the Flow state [5].

- Clear goals (expectations and rules are discernible and goals are attainable and align appropriately with one's skill set and abilities). Moreover, the challenge level and skill level should both be high.
- Concentrating, a high degree of concentration on a limited field of attention (a person engaged in the activity will have the opportunity to focus and to delve deeply into it).
- A loss of the feeling of self-consciousness, the merging of action and awareness.
- Distorted sense of time, one's subjective experience of time is altered.
- Direct and immediate feedback (successes and failures in the course of the activity are apparent, so that behavior can be adjusted as needed).
- Balance between ability level and challenge (the activity is neither too easy nor too difficult).
- A sense of personal control over the situation or activity.
- The activity is intrinsically rewarding, so there is an effortlessness of action.
- A lack of awareness of bodily needs (to the extent that one can reach a point of great hunger or fatigue without realizing it).
- Absorption into the activity, narrowing of the focus of awareness down to the activity itself, action and awareness merging.

B. Flow model for competitive type video game

In the case of video games or sports, the challenge level is replaced by the competitor's skill. In other words, a well balanced level between the competitor's skill and player's skill may induce Flow on the player. Fig. 2 shows estimated model of the mind state of the competitive type game with the competitor's skill as the vertical axis instead of the challenge level.

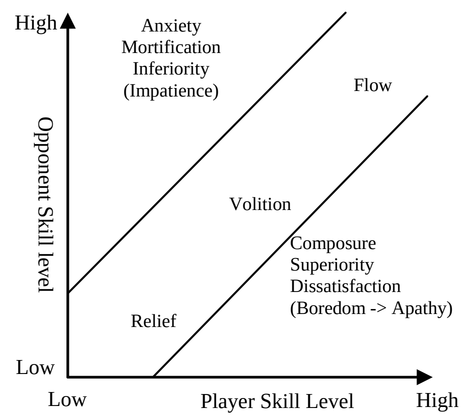


Figure 2. Proposed model of Flow state (Competitive type).

When the competitor's skill level is higher than the player's skill level, the player may feel 'Anxiety', 'Mortification' and 'Inferiority'. In the opposite case, he will feel 'Composure', 'Superiority' and 'Dissatisfaction'. When their skill levels are almost equal, the player will feel Flow and 'Volition'. Furthermore, player's emotions will change with time. The emotions which are indicated in '( )' show the player's feelings after the interval has passed.

C. Flow model for cooperative type video game

In the case of the cooperative type game, the challenge level can be replaced by the cooperator's skill. Fig. 3 shows the estimated model of the mind state in the cooperative type game with cooperator's skill level as the vertical axis. When the cooperator's skill level is higher than player's skill level, the player may feel 'Luck' and 'Impatience'. In the opposite case, he will feel 'Boredom'. When their skill levels are almost equal, the player will feel 'Volition', 'Relief' and Flow. The emotions which are indicated in '( )' also show the player's feelings after the interval has passed.

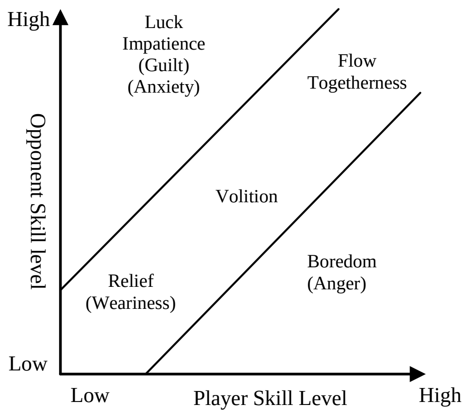


Figure 3. Proposed model of Flow state (Cooperative-type).

After each competition, the subject answers a questionnaire. The subjects are 15 male and female students in their twenties.

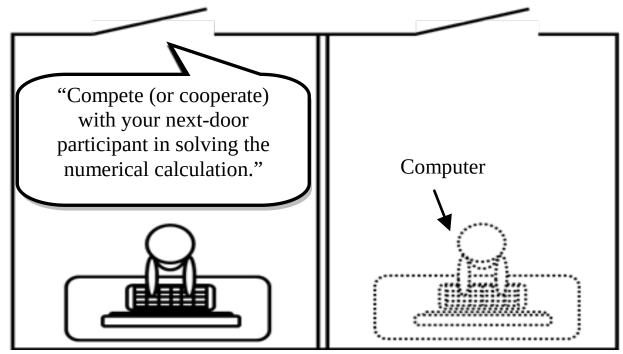


Figure 4. Experimental setup.

III. Influence of Relationship Between Player and Opponent on the Player's Performance and Emotion

A. Experimental method

To investigate the influence of the relationship between player and opponent on the player's performance and emotion, we performed some sensory evaluations under various conditions. For that purpose, a simple calculation game was produced. Using this system, the subjects solve a numerical calculation under controlled conditions; the opponent is a competitor or a cooperator and the opponent's skill level is high/even/low ability. The game is over when the 50 calculations are done by the subject and the opponent. After the game, we make subjects respond to a questionnaire. The subjects performed each of these conditions randomly.

B. Experiment I (Competitive type game)

First of all, the subject is asked to go into the room and to sit down on the chair in front of the PC shown in Fig. 4. Secondly, he is asked to solve some two-digit numerical calculations on the computer display for practice. After that, the experimenter lets the subject know that there is another subject in the next room, and he will compete with the competitor in solving the numerical calculations on the display. The victory will be determined by the number of solved calculations. In fact, the competitor is the computer and the competitor's skill is controlled based on the subject's skill, which is measured in the practice round. The competitor's skill (computer) is set to high ability, equal ability or low ability. It means the competitor's calculating speed is twice, even or half of the subject's calculating speed. In the experiment, the subject doesn't know the competitor's ability. Fig. 5 shows the G.U.I.: Graphical User Interface of experiment I.

The numerical calculation is displayed on the left side, and the numbers of solved calculation of the subject and competitor are shown in the lower region in number form and in a bar graph on the right side. The subject competes with 3 opponents, and the order of competition is random.

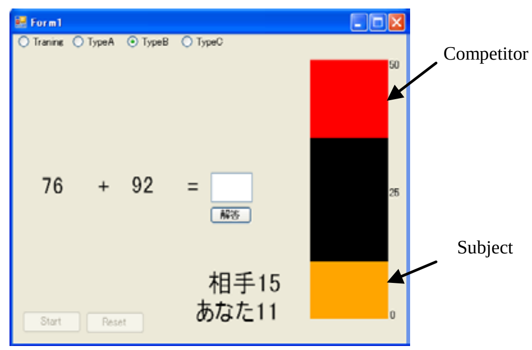


Figure 5. G.U.I. of experimental system (Competitive type).

C. Experiment II (Cooperative type game)

For Experiment II, a new set of 15 subjects were recruited. Experiment II is almost the same as I. However, the experimenter lets the subject know that there is another subject in the next room, and he will cooperate with them in solving the numerical calculations on the display. Fig. 6 shows the G.U.I. of experiment II. The numbers of solved calculation of subject and cooperator are again shown in the lower region in number form and in a bar graph on the right side.

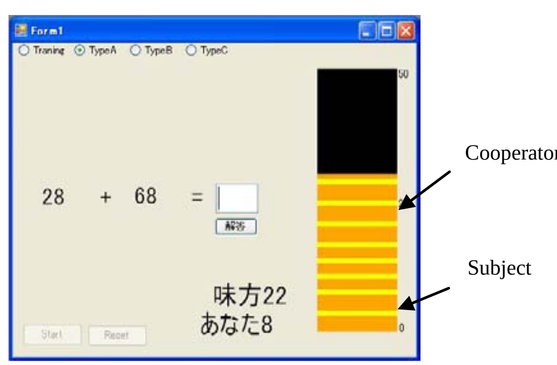


Figure 6. G..U.I. of experimental system(Cooperative type).

 ISBN: 978-1-61208-250-9

IV. EXPERIMENTAL RESULTS

After each experiment, the subject intuitively plotted the point which indicated the relationship between the subject's skill level and the opponent's skill level on an empty graph in which the horizontal axis shows the subject's skill level and the vertical axis shows the opponent's skill level. Fig. 7 and 8 show the accumulation of all subjects' perceived comparative skill level entries. In Fig. 7 and 8, the black circles indicate the high ability opponents; the squares indicate the equal ability opponents and triangles indicate the low ability opponents. From these figures, it is clear that most subjects recognized the opponent's ability exactly. Additionally, the white circles show the games that the subject thinks were the most interesting of the 3 levels. The star mark shows the center point of those preferred games. The subjects showed a tendency to prefer the game in which the opponent's skill level is almost the same as the subject's skill level in both conditions.

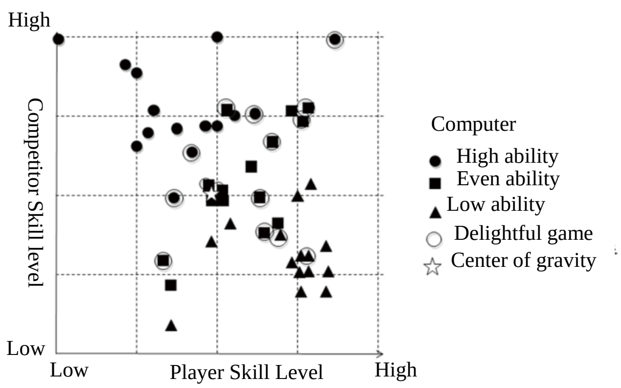


Figure 7. Questionnaire result (Competitive type).

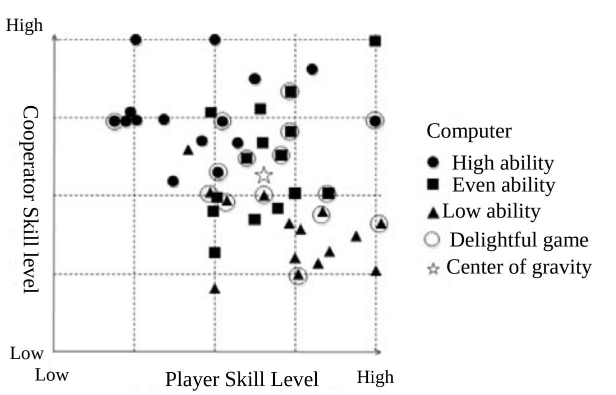


Figure 8. Questionnaire result (Cooperative type)

Fig. 9 and 10 show the average solution times per one numerical calculation of all subjects. In these figures, the gray bar shows the most interesting game and the black one shows the others. These data are normalized by the solving time recorded in practice. All average times are lower than those recorded in practice with the most interesting games receiving even lower times than the others. Especially, in the competitive format, there is a significant difference at 5%, which indicates that the player performs best when he feels enjoyment or feels best when he performs well in a competitive environment.

Fig. 11 and 12 show the emotions that the subjects felt in each game.

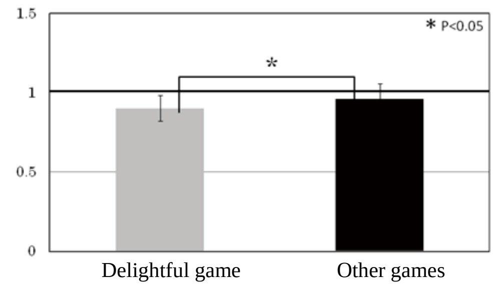


Figure 9. Average of response time (Competitive type).

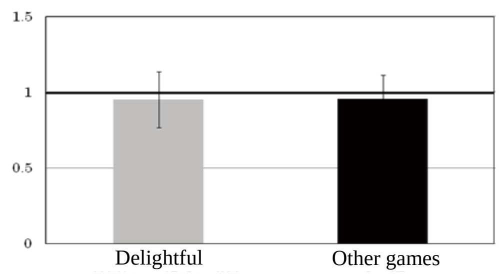


Figure 10. Average of response time (Cooperative type).

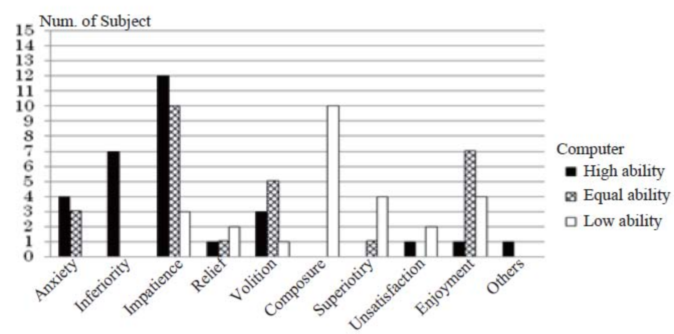


Figure 11. Questionnaire results for subject's feeling (Competitive type)

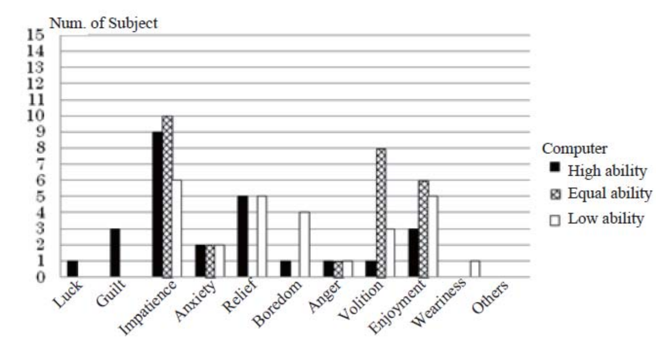


Figure 12. Questionnaire results for subject's feeling (Collaborative type)

In the case of the competitive game, the subjects tended to feel 'Composure', 'Superiority' and 'Enjoyment' when the opponents' skill levels were lower than theirs. In the case of the cooperative format, they felt 'Impatience', 'relief', 'Enjoyment' and 'Boredom'. When the opponents' skill level was higher than the subjects' skill level, they tended to feel 'Impatience', 'Inferiority' and 'Anxiety' in the case of the competitive type. In the case of the cooperative type, they felt 'Relief' and 'Enjoyment' in addition to 'Impatience' and 'Guilt'. When the opponents' skill levels were equal to the subjects' skill levels, they tended to feel 'Impatience', 'Volition' and 'Enjoyment' in both game types. From these results, it is possible to suggest that even the same game style can induce the various emotion states by controlling the situation.

Fig. 13 and 14 show the emotions which are the top 3 in Fig. 11 and 12, on a graph which indicates the player's skill level as the horizontal axis and the opponent's skill level as the vertical axis. Comparing this to Fig. 2, we obtained almost the same result shown in Fig. 13. However, when the ability of the opponent is lower than that of the player, we expected the emotion 'Dissatisfaction', yet there aren't so many subjects who selected this emotional state. In the case of the cooperative type game, when the opponent's ability is lower or higher than the player's ability, many subjects responded 'Relief' and when the opponent's ability was lower, they tended to feel 'Enjoyment' contrary to expectations.

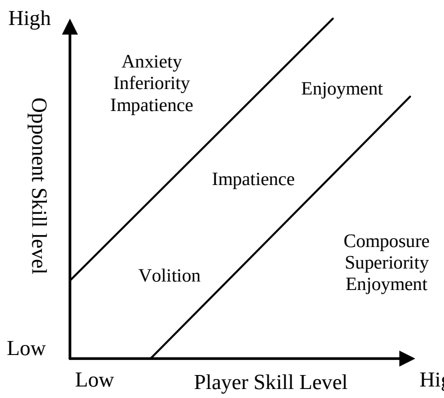


Figure 13. Flow model based on questionnaire result (Competitive type)

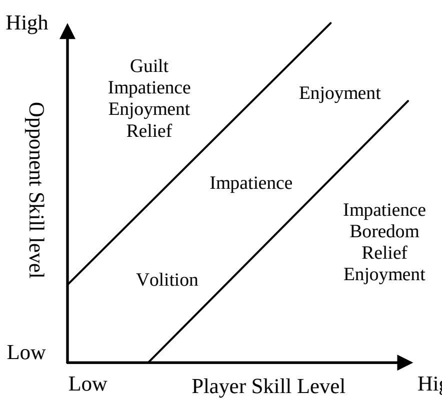


Figure 14. Flow model based on questionnaire result (Cooperative type).

The emotion 'Enjoyment' is chosen in almost all situations, except for the situation that the opponent's skill is higher than that of the subject's within the competitive type format. In the case of the cooperative type game, the subject tends to feel 'Enjoyment' and 'Impatience' in all situations. On the other hand, in the case of the competitive type, the difference emotions were induced by the opponent's ability. When the ability of the opponent is equal to the subject's ability in both cases, 'Enjoyment', 'Impatience' and 'Volition' are selected.

The results of the questionnaires based on the component of Flow are shown in Fig. 15 and 16, when the opponent's ability is equal to the subject's ability the highest score is obtained in both game types. They also indicated that the game player is able to enjoy playing most when the opponent's skill is equal to the subject's skill in any game. The situation in which the opponent's ability is higher than the player's ability ranks second in the case of the competitive type game, however it ranks lowest in the cooperative type game. From these figures, the condition when the opponent's ability is equal to the subject's ability can provide the sense of flow best to the subjects. Comparing these results with Fig.13 and 14, there is high possibility that 'Enjoyment', 'Impatience' and 'Volition' are selected as the emotional elements of flow.

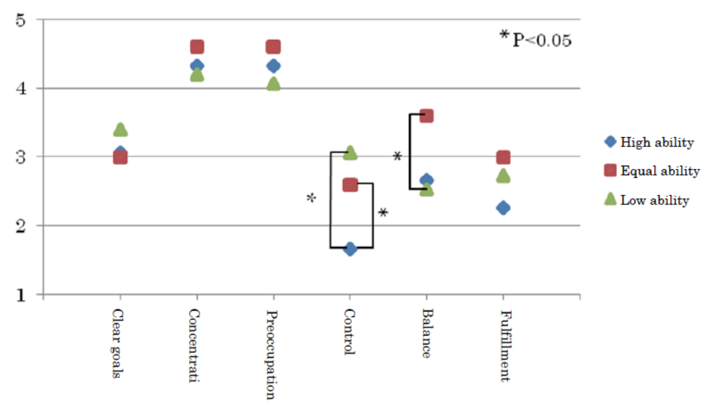


Figure 15. Questionnaire result to the flow component (Competitive type).

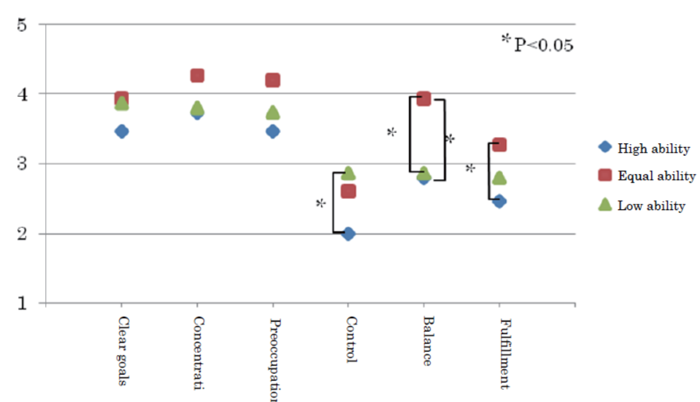


Figure 16. Questionnaire result to the flow component (Cooperative type).

If it is necessary for a game which provides the state of being crazy that the components of flow are high, that game has to provide the situation that creates the feelings not only 'Enjoyment' but also 'Volition' and 'Impatience' on the user's mind.

V. CONCLUSION AND FUTURE WORK

In this study, to investigate the influence of game types and opponent's ability on the player's state of mind and performance, we developed a simple numerical calculation game system. Through use of this system, we performed some experiments under controlled situations. All subjects solved the calculations at a similar rate, except when factoring for the players' recognition of the cooperator or opponents' skill level. However, the subjects' performance and emotions are different depending on the situation, which suggests that the video game system has an effect on the player's emotional state.

We have a plan to evaluate the influence on the player's skill in the game with team on the player's performance and state of mind.

On the other hand, the experiments in this study were performed under a controlled video game situation. However, we think that the results of the experiment may be applied on the many and varied systems which support human performance, emotion and motivation. We have another plan to construct the G.U.I. for operation system on the tablet pc and mobile phone, which will adapt the user's skill and usage based on these experimental results.

REFERENCES

[1] C. Mihaly, Flow: The Psychology of Optimal Experience, New York: Harper and Row, 1990.

[2] I. Kuramoto, K. Kashiwagi, T. Uemura, Y. Shibuya and Y. Tsujino. "Weekend Battle: an Entertainment System for Improving Workers' Motivation", Proc. of 2nd International Conference on Advances in Computer Entertainment Technology (ACE2005), pp. 43-50.

[3] S. Tamakoshi, T. Takahashi, M. Terao, D. Sawai, A. Imanishi, F. Morimoto, A. Yagi and H. Katayose, "A study in the entertainment of fighting video game measured by fNIRS : a comparison vs. human and vs. computer", IPSJ SIG Technical Report EC-3, (2006) in Japanese.

[4] J. Takatalo, J. Hakkinen, J. Kaistinen and G. Nyman, " Presence, Involvement, and Flow in Digital Games", Evaluating User Experience in Games, pp.23-46, Springer.

[5] D. Weibel and B. Wissmath, "Immersion in Computer Games; The Role of Spatial Presence and Flow", Int. J of Conputer Games Technology, Vol.2011, Article ID 282345, 14pages,2011.

[6] D. Weible, B. Wissmath, S. Habegger, Y Steiner and R. Groner, "Playing online games against conputer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment", Conputers in Human Behavior 24 (2004), pp.2274-2291

[7] C. Mihaly, Finding Flow: The Psychology of Engagement With Everyday Life. Basic Books, 1998.

Reducing the User Burden of Identity Management: A Prototype Based Case Study for a Social-Media Payment Application

Till Halbach Røssvoll, Lothar Fritsch
Norwegian Computing Center, Oslo
http://nr.no

***Abstract*—Payment applications inside social media dealing with privacy and security sensitive content require, besides trust in the involved parts like financial institutions and providers of electronic identities, in particular the trust of the users. The e-Me project focuses on this trust and aims at providing multimodal, adaptive authentication and authorization methods for social media that are usable for all users. In an integrated social-payment application connected to online banking, an OpenID provider has been developed by means of inclusive-identity management methods. The provider is used for both the social-media access control and the embedded payment service. This work describes the design decisions and eventual design made for the prototypes with considerations concerning both e-inclusion and information security and privacy.**

***Keywords*-Trust, security, privacy, identity management; e-inclusion, accessibility, usability, universal design; social media/networking applications**

I. INTRODUCTION

For architects and developers of the ever growing number of social media and electronic services, access and identity management (AIM) is a substantial part of the system design. AIM refers to techniques for determining and organizing the identity of a user in order to grant access to a service or data, also referred to as authentication, or to authorize the execution of a task [1]. In addition, electronic identities are used to organize personal data, and to provide advanced identity management systems [2].

In the context of information technology, accessibility describes the degree to which a solution is accessible for as many people as possible, in particular those with impairments, and those using assistive technology together with the product or service. Usability refers to the ease with which people can use a particular product or service. Obviously, AIM applications have to be both secure and privacy aware, and should at the same time be as accessible and usable as possible. Researchers have previously pointed out the need for inclusive access and identity management [3], as only few of these systems pay attention to accessibility and usability issues [4].

This work discusses how to design for trust, privacy, and security regarding online services and applications inside social media, while at the same time meeting the requirement for universal design. Examples of such applications are image galleries, music sharing services, online games, and news feed services. The term trust is used here in its most generic sense as the degree of reliance of one entity on another [5]. The definition of privacy in this work is based on the EU Data Protection directive [6], in the spirit of the data subject's informed participation, while security in this work bears the meaning "degree of protection to safeguard the asset, here personal data, against threats in terms of data exposure, damage, or loss".

Parts of this work have been presented on a previous occasion [7]. The novel contributions of this article are

1) an in-depth description of the prototypes, including privacy and security aspects,
2) a discussion of technical decisions regarding the universal design of the system,
3) the discussion of design considerations regarding the system's functionality for privacy and security, and
4) a discussion of the implications of the design for the user's experience of trust.

The work is organized as follows. First, the scope providing research project is introduced, followed by a description of the proof-of-concept application developed in the course of the project. Then, e-inclusion, trust, and privacy aspects are discussed, and a number of trust establishing measures is presented as a checklist before the paper concludes.

II. THE E-ME PROJECT

The research project e-Me sets the context for this work [8]. The main goal of the project is to provide new knowledge to improve the usability and accessibility of access and identity management systems, including authentication mechanisms, in social media without compromising privacy and security, and without offending legal frameworks.

In the course of the project, the example application PayShare has been developed. The starting point for the development of a highly usable and accessible prototype was a literature review on the field of accessibility and usability issues of personal identification systems [9], recommending —- among others:

- an open and universally designed solution with an accessible, adaptive, and personalized multimodal user interface,

 ISBN: 978-1-61208-250-9

- a minimally exposed user profile with reasonable defaults and opt-ins,
- and the application [1] of privacy-enhancing technology.

With this in mind, several hypotheses were set up:

1) The majority of users is suffering from having to handle too many user names and passwords for authentication.
2) The majority of current authentication mechanisms is not accessible to users with impairments.
3) Users have different requirements and preferences for privacy and security in electronic products.
4) Users experience multiple authentication processes in case of frequent authorization as cumbersome.
5) Authentication as used in social media can be applied to privacy and security aware applications without a degradation of the level of security or privacy.

In short, the solutions provided by PayShare are:

1) OpenID cuts down the numbers of service accounts to remember for the user.
2) Authentication adaptation by means of several OpenID login alternatives, namely password, a series of pictures, a series of sounds, pattern, and personal question, which — in total — have a higher degree of accessibility than just a single login method.
3) User defined threshold for the application of more frequent authentications.
4) Validity of a person's authentication for a user defined time span.
5) OpenID as an authentication means to authorize payments in an financial application inside a social medium.

All aspects of the solution are subsequently explained in more detail.

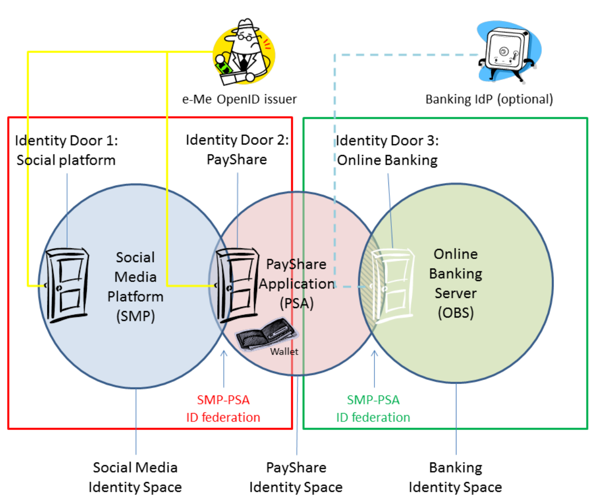


Figure 1. Identity federations and identity doors in the PayShare application

III. THE PAYSHARE APPLICATION

The proof-of-concept prototype PayShare is a means to test design principles, user interface, and system functionality. It can be described as an online payment service. Figure 2 shows the flow diagram of the PayShare application. The application's three entry points are "Add new claims", "View claims", and "View single claim". The user is automatically guided through the block sequence {"Read about", "View/accept terms"} before anything else can be done inside the application.

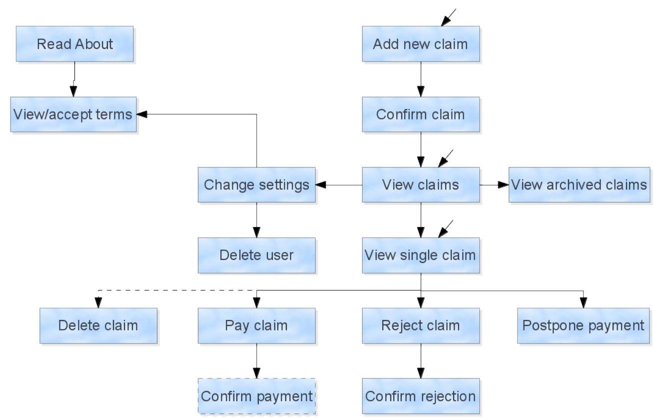


Figure 2. The flow in the PayShare application, which equals the application's pages/screens. The dashed arrow indicates a path that is only open to creditors. The dashed box refers to the external OpenID provider.

Upon registration, which basically requires the user's acceptance of the Terms&Conditions of the service, users can file payment claims for the entire group, assuming all group members are also connected through a particular social medium. For instance, consider a group of friends out to travel. One of them, here called creditor, pays for the travel tickets of the entire group and files claims against all "friends", here referred to as debtors, in PayShare. The debtors then get notifications that there are open claims that they have to pay, and the creditor can conveniently track any payment progress, i.e., see who has paid or not. Payments can be made directly in the PayShare service, meaning inside the social medium, to/from a virtual wallet in form of so-called credits, or to/from an account in a trusting bank. There is the possibility to transfer money from the bank account to the virtual wallet. PayShare in its entirety meets the requirement represented by the fifth hypothesis (Section II).

Figure 3 shows the screen shot of the "View single claim" block which is central to the application. It consists of two logical units. One unit gives all necessary claim details, such as amount, creditor, explanatory message, date, and the subclaims for all the other group members. Strictly speaking, the latter information is not mandatory for the debtor to know in order to pay the claim, but it serves as an additional explanation and justification of the claim. In the logical unit

 ISBN: 978-1-61208-250-9

below the claim details, the debtor is asked for her choice; that is, pay the claim from the wallet, pay the claim from the bank account, reject the claim, or postpone the payment. Concerning the latter, the claim will be maintained, and a reminder about it is send to the debtor from time to the other. As mentioned, the debtor could also reject the claim, with or without a personal comment to the creditor. Before the navigation buttons are shown on the page, a brief message is displayed informing the debtor about the payment; i.e., if (and why) an additional authentication might be necessary, or — if applicable — just informing the debtor that the payment is carried out without authentication. A hyperlink to the appropriate part in the settings is given where these security preferences can be changed.

A. *Privacy & security aspects*

To solve security and privacy challenges, PayShare aims at the separation of identity spaces from each other, see Figure 1. To achieve this, electronic identities and their use have been organized in Identity Spaces. These spaces are organized with PayShare as a intermediary separating the identity spaces. We use the concept of an "Identity Door" to describe the handling of e-ID when authenticating to a service or when authorizing a transaction. In Figure 1, we show how three such identity doors are defined in PayShare:

- Door 1 is the entrance door to the social media platform that will provide the social context for PayShare.
- Door 2 is the registration to PayShare, which in general is entered from the social media platform (but can be accessed through its own user interface).
- Door 3 is the additional authorization and interface against the online banking transaction server.

Since security usability — especially of authentication mechanisms — is the primary objective of the e-Me project, compromises between the security infrastructure, privacy design, and usability had to be taken. These will be discussed further below.

Regarding other security ensuring measures, there is a user-defined threshold for the payment amount. If the amount is above the threshold, an additional OpenID authentication of the user is required to authorize payments, as illustrated in Figure 1 with Door 3. However, as the authentication mechanisms used by banks are numerous, and as an extra authentication door is perceived as a hindrance by sensitive users, PayShare acts rather as a payment intermediary. This approach allowed us to restrict Door 3 authentication to cases where the wallet is charged or discharged. As such, users are enabled to put a price at the convenience of not having to go through an authentication process for small-amount payments. Payments with an amount below the threshold are one-click payments. This measure meets the requirement represented by the third hypothesis Section II. At the same time, the OpenID provider can offer user friendly and inclusive authentication methods both on Door 1 and Door 2.

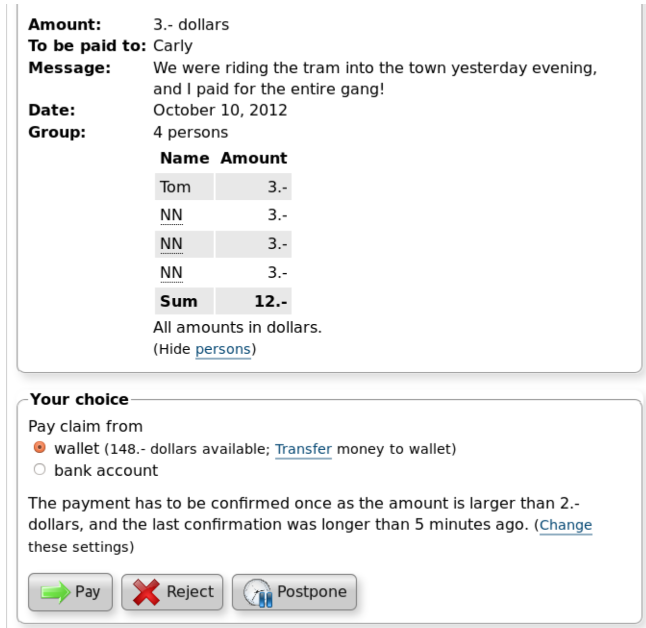


Figure 3. Screenshot of the "View single claim" page in PayShare.

Authentication is delivered by an OpenID provider, which has the advantage that only a single password has to be remembered by the user, even for a variety of applications, authentication contexts, and web sites. This measure meets the requirement as formulated by the first hypothesis Section II. Theoretically, the solution works together with any OpenID provider with full support of the OpenID protocol, but in practice only a few selected providers may be found suited due to trust reasons. An own OpenID provider has been developed as one of the deliverables of the e-Me project.

Data from any authentication, be it login to the social medium or authorization of a payment, expires according to a user-defined time span for authentication validity. This measure meets the requirement put forward by the fourth hypothesis (Section II) and aims at avoiding frequent cumbersome authentications as the user's ID is remembered for this specific time span. Consequently, an authentication process is only invoked when the payment's amount is above the threshold, and when the time elapsed since the last authentication has exceeded the authentication validity duration. More on the reasoning for the respective design choices in Section IV-B.

IV. Discussion

In the following, it is discussed how accessibility and usability issues as well as trust, security, and privacy matters are reflected in the PayShare application.

 ISBN: 978-1-61208-250-9

A. e-Inclusion

As mentioned before, the project's focus is on inclusion aspects of the solution. Its target groups consist of users with various impairments, and the elderly. Acknowledged impairments are cognitive challenges such as dyslexia and dyscalculi, orientation, and memory problems, sensory challenges like vision and hearing reduction, and motor challenges like trembling hands. Elderly users are likely to have a combination of impairments. However, apart from these focus groups, PayShare is required to be universally designed, meaning that it can be used by virtually all persons.

To meet the requirement as formulated by the second hypothesis (Section II), the aforementioned OpenID solution offers in total five login alternatives:

- Password memorizing,
- recognition a series of pictures,
- recognition of a series of sounds,
- pattern drawing, and
- knowing the answer to a personal question.

What login method is used is up to the user to decide. Each of the alternatives aims at a particular target group (but is, of course, not restricted in use to this group): Password for user without particular login problems, picture series for the hearing impaired, sound series for the vision impaired, pattern drawing for dyslectic individuals, and personal question for users with short-/medium-term memory deficits. In total, e-Me authentication has a higher degree of accessibility than just a single login method.

Apart from login, the system's inclusiveness is met by a number of measures concerning universal design. For instance, the user interface is tailored to the needs, preferences, and context of the respective user by means of a user profile, satisfying major parts of the first requirement from the literature review.

B. Privacy implications

The PayShare application accounts for privacy concerns in a number of ways.

The service follows the privacy requirements [6] derived from the EU Data Protection Directive [10]. In addition, it considered the potentially harmful actions laid forth by Solove [11], and places particular attention on the handling of electronic identifiers according to the PETweb II risk taxonomy [12]. According to Solove, harmful actions on personal data are:

- Information Collection: Collection and accumulation of personal information can cause harm.
- Information Processing: Handling of collected personal data that can cause harm.
- Information Dissemination: Harms of spreading, or threat of spreading information.
- Invasion: Risks of intrusion and decision interference.

The following requirements from the uTRUSTit privacy requirement report [12] were used as the essential requirements following from the EU data protection directive [10].

- Personal data is only processed after the person gave informed consent for the processing.
- A person has the right to inquire about the own personal data stored with another party.
- A person can revoke consent for personal data processing given earlier at any time.
- The data processor can only process personal data according to the given consent (e.g. according to a privacy policy).
- The data processor makes sure that personal data is sufficiently protected from unauthorized access, manipulation, abuse or loss.
- The data processor makes sure to be able to react accordingly to consent revocation for any given personal data.

The PETweb II risk taxonomy for electronic identity management systems found a number of factors that constitute privacy risk in handling identifiers [11]. Major risks found include the possibility for identifiers to get profiled, to get stolen, and to get remotely used. In the PayShare scenario, Door1 authentication leads the users into a personalized profile. The purpose of authenticating is to access the same profile every time. The same holds for PayShare's Door 2 authentications , where personal money is administered. The risks for stolen identifiers is strongly dependent on the authentication technologies offered by the OpenID provider. In e-Me, experiments are performed with easy-to-use password replacements, and with signature strength smart card solutions [13].

A number of compromises have been made. In the given scenario – payment among a group of users registered to a social media platform, in connection to their bank account – full anonymity was nowhere possible. The social media profile, even if it should be created under pseudonym, gets through the PayShare service connected to a bank account that is connected to a verified person. In addition, pseudonymity on the social media is of limited usefulness in a scenario where friends split the dinner bill – as the identities are known between the friends anyway. The critical pieces of personal information were identified to relate to the occurrence, frequencies and destinations of payments. By inclusion into a social media, PayShare is principally able to trace a friends network's payment behavior. Repeated payments between friends can indicate closer relationships, reveal business secrets, or prove social relationships to others. The introduction of the Wallet for low-value payments loosens the connection between the bank and the social media activities. By acting as a "wallet filled with small change" up to the set threshold, the bank gets only involved into higher-value transactions. However, as ultimately, bank transactions will be used beyond Door 3, there is no way around compliance with financial market regulation, and thus full identification of sender and receiver in a bank transaction.

Regarding concrete privacy measures, there are several hyperlinks to the service's Terms such that the user can view this document also after its acceptance upon registration. In case the user does not agree with the Terms anymore, an option to delete the user account is provided. There is also a hyperlink from the settings to the deletion of the user account, as the user might find the security and privacy settings inappropriate for own needs.

As mentioned in Section III, the view of a single claim contains the display of the subclaims for group members. However, all names (except the own one, of course) are anonymized per default to honor the privacy of all involved persons. The visibility of claims can be set to "Visible to all group members" in the settings, if desired.

The deletion of a user's account includes the irreversible removal of the user's profile which in turn includes settings like bank account number, visibility of claims, OpenID address, threshold for additional authentication, and duration for storage of authentication realm. Additionally, the user's ID is removed from all open and archived claims in the database, following this policy: Where the user is a creditor and the (sub-)claim is open, the (sub-)claim is deleted. In archived (sub-)claims with the user as a creditor, the user ID is erased from the database entry. Other claims, in particular those where the respective user has been named as a debtor, are not touched.

Creditors may delete own claims, but here the restriction is that only open claims may be deleted, while archived claims cannot. In this case and in the instance of the user account deletion, the privacy of the creditor and the privacy of the debtor have to be weighted against each other. PayShare accounts for the debtor's privacy by keeping archived records, but at the same time also honors the creditor's privacy and erases her user ID from the record, as this is information provided by the creditor herself. A debtor cannot, however, delete claim records as the information concerning who is the claim's debtor has been provided by the creditor and is hence out of control for the debtor.

C. Trust implications

As a payment service that voluntarily relies on OpenID for authentication, and that is linked to a bank account, there are a number of trust chains. Figure 4 illustrates all paths of trust.

PayShare acts as an identity broker against the social media and the bank identity space. The bank (the trustor) has, as a minimum requirement, to trust PayShare (the trustee)

- that all virtual credits are safe and properly handled with the service,
- that all payments are correctly executed, and
- that personal data are handled in a confidential way.

The same applies to the trust chain from the user (the trustor) to PayShare (the trustee). Needless to say, the user has to trust the bank in the same way as she trusts PayShare. In addition, the user needs to trust the social medium, which provides the framework for services like PayShare, for data that PayShare and the medium have in common, like user names, and the OpenID provider, for handling authentications in a secure and privacy aware way. PayShare needs to trust the underlying social medium as well as the OpenID provider for proper and secure authentication.

Besides the robustness of the OpenID provider's authentication methods, PayShare itself must be seen as the most critical point in the privacy infrastructure. PayShare is the entity that collects and combines both information about the users' social media identity, their bank identity, and their payment requests. Therefore is the correct handling and deletion of payment requests and payment information at PayShare of utter importance. The issue of the trusted 3rd party is a well-known problem for privacy and trust research. Often, for instance when personal data is part of a business model, some form of 3rd party (e.g. intermediaries and brokers) are needed to perform the business function. These brokers can help to protect privacy by separating identity spaces and personal data between the participating stakeholders. The concept has successfully been used to anonymize location data in location-based services [14], [15]. In PayShare, the PayShare service itself handles the identity spaces with a direct mapping rather than using anonymizers [16] and advanced attribute credentials [17] due to the fact that banking transactions cannot legally be anonymized.

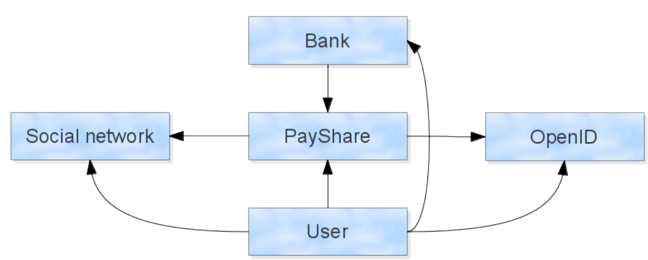


Figure 4. Chains of trust among the PayShare stakeholders.

Is is argued here that especially impaired users need particular measures to develop the same degree of trust like the average user to make up for the impairment. For instance, a blind person who is enabled to use an audio-based authentication scheme will have more trust in such a system that accounts for the user's preferences than systems that only offer visual CAPTCHAs. Similarly, a person with orientation and problem solving problems is likely to develop more trust in a system showing explaining messages about what

 ISBN: 978-1-61208-250-9

is about to happen on the current screen than systems that do not have the same degree of usability.

All users build their trust on pieces of information, about what is about to happen, why certain things happen, what the user's choices are, etc. The picture is never entirely complete; rather, the more information the user has at a given point in time, the better in order to be able to carry out a particular task. Impaired users are likely not to get as much information as ordinary users, except when the solution is universally designed and thus accessible to virtually all.

On the other hand, too much information can result in the disorientation of the user, especially those with orientation and learning challenges [18]. The solution should therefore limit the amount of information the user is confronted with, and should also ease its processing, in terms of a weighting of the information's importance for the context the user currently is in. To sum up, it is crucial in particular for impaired users — but also for all users in general — that the system provides and makes accessible all the information the user needs in order to carry out a task by herself, not less but not more either to avoid user confusion.

V. Conclusion

User control and information are crucial to achieve a high degree of trust of the user to the service. There is a connection between e-inclusion and trust in terms of the fact that a high degree of accessibility and usability empowers the user in certain situations to use the respective service at all. In other situations, it increases the user's control or feeling of control and thereby the user's trust. The perception of increased trust is not only applicable to users with impairments but rather all users, as it is widely recognized that e-inclusion measures for particular focus groups generally increase the service's usability for everybody.

A particular challenge in the realization of PayShare was its connection to several identity spaces. Social media identities and online banking identities are different in scope, robustness, and use. PayShare had to accept the role as an identity broker or identity intermediary to bridge the respective identity spaces. The resulting privacy issues were analyzed and led to the above requirement list for PayShare.

Acknowledgment

The e-Me project is funded by the Norwegian Research Council under project no. 201554.

References

[1] K. S. Fuglerud and T. H. Røssvoll, "Usability and accessibility of personal identification management systems in electronic services," in *Proceedings of eChallenges-2011*. Florence (Italy): IIMC International Information Management Corporation Ltd., Oct. 2011. [Online]. Available: http://echallenges.org/e2011/default.asp?page=paper-repository&fltyear=all&flttheme=all&flttype=all&flttitle=&fltauthor=halbach&pagesize=100&submit=Search

[2] WP3, "Fidis Deliverable D3.1: Structured overview on prototypes and concepts of identity management systems," FIDIS Consortium, Tech. Rep., September 2005.

[3] L. Fritsch, K. S. Fuglerud, and I. Solheim, "Towards inclusive identity management," *Identity in the Information Society*, vol. 3, no. 3, pp. 515–538, 2010.

[4] G. Sauer, J. Holman, J. Lazar, H. Hochheiser, and J. Feng, "Accessible privacy and security: a universally usable human-interaction proof tool," *Universal Access in the Information Society*, vol. 9, no. 3, pp. 239–248, 2010. [Online]. Available: http://www.springerlink.com/content/j801373hw352514r/

[5] L. Fritsch, A.-K. Groven, and T. Schulz, *On the Internet of Things, Trust is Relative*, ser. Communications in Computer and Information Science. Amsterdam: Springer, 2011, vol. 277, ch. 9, pp. 267–273.

[6] European Council and Parliament, "Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on Privacy and Electronic Communications)," 2002.

[7] T. H. Røssvoll, "Trust implications for universal design of social-networking applications," User-Centered Trust in Interactive Systems Workshop at NordiCHI 2012, Oct. 2012. [Online]. Available: http://www.nordichi2012.org

[8] Norwegian Computing Center, "e-Me — Inclusive identity management in new social media," 2012. [Online]. Available: http://nr.no/pages/dart/project_flyer_e-me

[9] K. S. Fuglerud and T. H. Røssvoll, "Previous and related research on usability and accessibility issues of personal identification management systems," Norwegian Computing Center, Oslo (Norway), Tech. Rep. DART/10/10, Oct. 2010. [Online]. Available: http://publ.nr.no/5371

[10] L. Fritsch, "Social media, e-ID and privacy - background for the e-Me project," Norsk Regnesentral, Tech. Rep. DART/02/2011, April 2011.

[11] D. Solove, "A taxonomy of privacy," *University of Pennsylvania Law Review*, vol. 154, no. 3, pp. 477–564, 2006.

[12] E. Paintsil and L. Fritsch, *A Taxonomy of Privacy and Security Risks Contributing Factors*, ser. IFIP Advances in Information and Communication Technology. Springer Boston, 2011, vol. 352, pp. 52–63.

[13] L. Fritsch, "Privacy visualization requirements in the Internet of Things - a uTRUSTit FP7 ICT project note," Norwegian Computing Center, Tech. Rep., September 2012.

[14] ——, "Utprøving av Buypass e-ID og Altinn.no - resultat av smartkort-studie i e-Me-prosjekt," Norwegian Computing Center, Tech. Rep. NR Note DART/03/2011, April 2011.

[15] T. Koelsch, L. Fritsch, M. Kohlweiss, and D. Kesdogan, *Privacy for Profitable Location Based Services*. Boppard: Springer, 2005, vol. 3450, pp. 164–179.

[16] J. Zibuschka, L. Fritsch, M. Radmacher, T. Scherner, and K. Rannenberg, “Privacy-friendly LBS: A prototype-supported case study,” in *13th Americas Conference on Information Systems (AMCIS)*, P. o. t. A. C. o. I. Systems, Ed., 2007.

[17] D. Chaum, “Untraceable electronic mail, return addresses, and digital pseudonyms,” *Communications of the Acm*, vol. 4, no. 2, pp. 84–88, 1981.

[18] T. Halbach, “Towards cognitively accessible web pages,” in *Proceedings of Third International Conferences on Advances in Computer-Human Interactions (ACHI-2010)*, International Academy, Research, and Industry Association (IARIA). St. Maarten (Netherlands Antilles): IEEE Computer Society, Feb. 2010, http://www.iaria.org/conferences2010/ACHI10.html. [Online]. Available: http://publ.nr.no/5236

 ISBN: 978-1-61208-250-9

Heads Up: Using Cognitive Mapping to Develop a Baseline Description for Urban Visualization

Ginette Wessel
University of California
Berkeley, CA, USA
gwessel@berkeley.edu

Elizabeth Unruh
University of North Carolina
Charlotte, NC, USA
eaunruh@gmail.com

Eric Sauda
University of North Carolina
Charlotte, NC, USA
EricSauda@uncc.edu

***Abstract*— Kevin Lynch's work on urban legibility has taken on new importance as the delivery of information about cities has shifted largely to mobile computing devices. This study extends his work with the aim of quantifying the number and type of elements that constitute a competent cognitive map of a city. We conducted a user study of 109 student sketch maps of Chicago that test the frequency and nature of the elements identified by Lynch (path, edge, district, node and landmark), their interrelationship and the effect of gender, prior experience and scale. We find that (1) participants identify two distinct urban scales, one at the neighborhood level and the other citywide, (2) competent cognitive maps involve relatively small numbers of elements: 15 (+/-7), (3) the selection of elements for the sketch map may include any of the elements identified by Lynch, but the frequency of landmarks and districts is negatively correlated, (4) participants recall significantly more districts and nodes at the citywide level, and (5) in addition to Lynch's identification of physical landmarks, participants also identify landmarks by function; such functional landmarks are more frequent at the neighborhood level.**

Keywords-legibility; cognitive mapping; urban visualization.

I. Introduction

The image of the city, a phrase coined by Kevin Lynch in 1960, has been widely used by urban planners and geographers as a means of conceptualizing the way users perceive and organize spatial information. The five categories postulated by Lynch (path, edge, node, district and landmark) have been the basis of subsequent research. Despite this literature, there are two salient research questions that remain unanswered:

- First is the epistemological question: are all the items in this set of categories at the same level of specificity; are they commensurable? Are there significant differences between the kinds of information that each category is describing?
- Second is the methodological question: is there a way to quantify the sufficiency of the elements and to test their relative frequency? In other words, are all these elements recognized by users as equally valid and are there significant relationships between their frequency of use?

The results of our empirical study will assist in the design of urban visualizations at all scales. We do so by uncovering those specific elements of the environment that may be regarded as most useful and recognizable. Specifically, our work is centered upon generating a baseline description of the urban environment from humans' cognitive maps.

Literature from urbanism, cognitive science and geography is a starting point for understanding environmental cognition. The seminal figure is Kevin Lynch who examines perceptions of urban environments in order to better understand how humans make sense of their surroundings. We adopt Lynch's concept of legibility defined as "visibility in a heightened sense, where objects are not only able to be seen, but are presented sharply and intensely to the senses" [1]. Lynch discovered that the mental maps which people use to help them find their way through cities are composed of five essential elements: paths, edges, districts, nodes and landmarks. Much like Lynch, we assert that legibility is crucial to the urban setting and involves visual and mental processes.

The central concept of such work is the idea of a spatial mental model [2], referred to by Lynch as mental map [1], or more generally understood as a cognitive map which proposes that the brain can acquire, code, store, recall and decode information about the relative locations and attributes of various phenomena in a spatial context. The development of a cognitive map often operates using schematization; some elements are emphasized and others are deemphasized to reduce cognitive workload. Cognitive maps of the environment cannot preserve metric information about the environment, but do preserve topological relations coherently. For instance, during wayfinding tasks it is common that a user's topological knowledge is more pertinent than knowledge of one's exact distance to an object and the precise dimensioning betweens objects. Therefore, one is more likely to be concerned with the type of elements recognized and remembered rather than the elements precise distance from one's own position in space. For this reason, we find hand-drawn sketch maps, a method commonly used by Lynch [1] and Appleyard [3,4], most useful as a means to capture schematized and topological knowledge about the environment.

In order to generate a baseline description of a cognitive map, we conducted a user study of architectural students after a field trip to the city of Chicago, distant from their home university. We asked each student to draw a sketch map of Chicago using an open question protocol. The

 ISBN: 978-1-61208-250-9

resulting sketch maps were coded to identify spatial characteristics of the city using paths, edges, districts, nodes and landmarks.

Our subsequent analysis was aimed at quantifying spatial elements in the sketch maps. How many of Lynch's elements did participants on average feel were necessary to use in order to give a clear depiction of the city? Just as important, we wanted to determine if the use of one type of element substitutes for another. If you identify a lot of landmarks, do you in general use fewer paths? Does previous knowledge of a city affect your use of elements? Does gender?

Our objectives in this study are:

- By means of a user study, extract a simple but complete set of spatial elements that are essential to constructing cognitive maps;
- Establish the base level of each element for a competent description of a cognitive map;
- Measure and study the relative interdependence of each element in describing an urban environment;
- Discover differences, if any, in the number of elements for users based on gender, previous experience and map scale.

The result of this study is an understanding of the number, type and interdependence of elements that form a cognitively adequate description of an urban setting that can serve as a guide in urban visualization design.

II. RELATED WORK

A. *Urban Visualization*

In recent years, several innovative mapping applications have used cognitive mapping, although without any quantitative understanding about the elements that constitute such a mental map.

LineDrive [5] is a navigation interface that incorporates principles of map distortion with the rendering of routes, adjusting lengths and angles to more closely approximate the kind of simplifications that mimic users experience. This approach focuses almost entirely on segment elements of routes, and demonstrates a subtle and effective adjustment to individual users. The shortcoming of this approach is the lack of contextual information, leading to problems with users becoming confused after a single wrong turn.

Copernicus [6] attempts to correct flaws within the LineDrive system by adding spatial context to the interface. Generally, this involves adding paths and neighborhoods (cities and towns) to the established routes from LineDrive. This represents an advance in terms of legibility, but leaves unsolved the issue of how much information to display and which types of elements should be used.

Researchers at Carnegie Mellon on the MOVE [7] system have developed mapping software that uses some of the principles from LineDrive in a two dimensional interactive network that adjusts as the user moves through the city. This corrects some of the original problems, but does not address directly the issues of cognitive mapping.

Some interfaces have attempted to incorporate landmarks into maps [8], focusing on selection processes for identifying landmarks and geo-referencing them on tourist maps. This represents an advance in the sort of data related to landmarks, but it does not address their relationship to the user.

There has also been some effort related to GPS navigation systems in cars, notably at SIGCHI 2008. These studies evaluate user behavior while in a car [9], performance of a driving simulator to study accuracy rates of proposed systems [10], and quick search versus categorical semantic search techniques [11]. But again, none of this research establishes guidelines for display visualizations.

Therefore, we identify a need to rely on the research of urban planners, cognitive scientists and geographers who examine the question of urban legibility. More specifically, we are interested in research that identifies a base set of spatial elements that are recognizable in any city. While each of the following fields of research focus on cognitive maps of the city in unique ways, we want to address the most relevant work in relation to our focus on element recognition, setting aside issues of spatial proximity judgments and element recognition sequencing.

B. *Cognitive Science*

To begin our overview of cognitive science related work, we should be clear about our use of the terms cognitive map and sketch map. By cognitive map, we mean the internal mental image that enables people to code, store and decode spatial information. By sketch map, we mean an external representation of a cognitive map that is solicited by the need to communicate in daily life (and by researchers).

The term cognitive map was first used by psychologist Edward Tolman in "Cognitive Maps in Rats and Men" [12]. He describes a maze previously mastered by rats that is blocked at a critical point and replaced by a series of radially arranged alternatives. His finding is that the rats greatly preferred the route that demonstrated an understanding of the spatial overview of the maze. Partly a reaction against strict behaviorism, his work led to the development of cognitive psychology.

A closer examination of cognitive maps later carried out by Ronald Briggs explains what factors cause their development [13]. He identified three complementary ways in which cognitive maps are created about a city: through an individual's sensory modalities, from symbolic representations such as maps, and from ideas about the environment that are inferred from experiences in other similar spatial locations. Much like Lynch's notions of legibility, Briggs also asserts that an individual's sensory modalities, including visual, auditory, olfactory and kinesthetic, play a significant role in cognitive map development. We find this insight useful in recognizing that cognitive map development is highly complex and involves more than just visual qualities. This work also explains the subjectivity and variability in such work.

C. *Urban Design and Planning*

As previously mentioned, the work of Lynch is a starting point for our study [1]. In addition to his claim that the city's image is represented in memory through five common elements, Lynch also develops three principle parts necessary for an environmental image: identity, structure and meaning. We recognize or identify objects, we notice a recognizable pattern, and we draw emotional value in relation to them. These principle components involved in constructing an image of a city are fundamental to our work and occur at the beginning of cognitive map development.

As an extension of Lynch's work on image construction, urban designer and environmental behaviorist Amos Rapport highlights the importance of meaning in the built environment by recognizing three levels of meaning [14]. Denotative meaning coincides with object recognition; a lower level meaning that manifests in identifying intended uses of a setting. A middle-level meaning, also called connotative meaning, refers to the emotional values associated with the object and is centered upon evaluative judgments such as how much you like the appearance of an area. Last, abstract meaning refers less to an object than to broader values. For instance, when looking at a place through "cosmologies, cultural schemata, worldviews, philosophical systems and the sacred," one experiences abstract meanings. Rapport's levels of meaning solidify the notion that legibility of the environment can be highly complex yet can be clearly organized based on specified personal values. Most importantly, people apply different types of meaning with different levels of significance to the same elements of the environment.

Donald Appleyard, a collaborator with Lynch, also worked in the field of environmental cognition and planning. While most of Appleyard's work focused on a view of the city from a navigational standpoint, his work in Ciudad Guayana [3] addresses the image of the environment as a tool to plan for a better community. He was interested in how residents structured their city and asked them to draw maps of their local area. Using these maps, he developed a categorization of various types of sketch maps (sequential, spatial, fragmented, scattered, chain, mosaic, branch and loop, and patterned). This method defines how residents conceptualize and structurally organize their city and shows that a recognizable path system is the main structural organizer for residents. Appleyard's study proves influential to our work in stressing methods of abstracting schematized information (i.e. topographical elements) from sketch maps.

D. *Geography*

The approach of geographers is not distant from the aforementioned theorists. As Peter Gould explains geographers who reach across traditional disciplinary boundaries to other social and behavioral sciences, find the truly satisfying explanation they seek to come from emphasizing the human as much as the geography [15]. For the following scholars, cognitive mapping proves a way to understand the spatial aspects of human behavior.

Geographer Reginald Golledge [16] focuses on the development of cognitive maps starting with specific landmarks to larger general areas. First, a person acquires declarative knowledge of discrete places, things and events. Next, they develop an understanding of a node and path sequencing of the environment. This provides the subject with a connective structure of transit paths and concentrated locations. Last, a completely integrated spatial representation is developed including characteristics of distance, direction, orientation, proximity, clustering and hierarchical ordering.

Similar to Golledge's work, findings from a study using participant sketch maps drawn over a period of ten months shows that landmarks and relative locations are among the first components that are learned, followed by paths, and then building from the framework of paths the initial relative locations become more precise [17]. As a three-step process, this serves as an incremental approach for which to understand the relationships and integration of elements in the environment.

Hintzman and others [18] reconfirm a sequential development of a cognitive map based on studies of orientation and target domains. They argue that instead of recalling the environment as a holistic cognitive map participants first recall the origin and target in memory, activate the shortest route between them and then span the route for a correct response.

III. User Experiment

A. *Procedure*

We conducted a user study of 109 architectural students after a four-day field trip to the city of Chicago. Within two weeks of returning from the trip, participants gathered in a study room equipped with a desk, a questionnaire and a writing utensil. To start, participants were given a brief overview of the study for the first time, but were not told about specific elements later used for analysis. Participants were then instructed to complete all questions in a thorough manner and to take as much time as needed.

B. *Experiment Questionnaire*

The experiment questionnaire primarily consisted of a recall exercise that included drawing a sketch map of the city. Other questions included background information about the participant such as gender, age and familiarity with the city.

The sketch mapping statement was phrased, "Now that you have traveled to Chicago, think about the environment you experienced. Imagine that you have a friend who has never been there before. Draw them a map to help them to understand and navigate Chicago." We intentionally left the statement open ended to allow the participant to draw freely in order to capture the most influential elements in each person's mental map.

C. *Participants*

Sixty-one males and 48 females participated in the study. Participants' ages ranged from 19 to 38 years old with an average age of 21. Based on questionnaire responses, 26 of the 109 participants said they had visited Chicago prior to the trip. Therefore, to ensure the results of the study were consistent, we analyzed the 26 sketch maps by testing for the equality of group means (student's t-tests) and found no statistically significant difference in any measure of element recognition in comparison to the participants who never visited Chicago prior to the study (Table I).

D. Data Analysis Procedures

Upon completion of the user study, we created a set of instructions for coding the maps based on Lynch's five elements. The instructions included definitions for each of the elements based on Lynch's description and a comparable example in a city other than Chicago.

In order to validate the clarity of the instructions, ten coders (4 professors, 6 students) were given the coding instructions and a random sampling of 30 maps. The coders performed content analysis by counting the number of each kind of element on the maps: paths, edges, districts, nodes, landmarks for their physical presence and landmarks based on their functional significance.

Researchers then studied the results of the ten coders and analyzed the consistency of coders' findings. Discrepancies with high significance were identified and consequently, researchers modified the coder's instructions to clarify the related areas. Researchers also determined that 13 of the 30 maps had a higher occurrence of discrepancies than the other 12 maps. These 13 maps, along with the revised instructions, were returned to the 10 coders for recoding.

The second round of coding with the revised instructions decreased the discrepancies between coder results. Next, two members of the research team coded all 109 maps per the revised set of instructions. The researchers compiled their results and determined a final set of element counts for the set of maps. The researchers also recorded the area in square miles for each map based on the elements noted as farthest from the city center using an online map digitizer tool.

IV. Results

Our analysis showed gender and prior experience visiting Chicago did not have a relationship to the total number or type of elements drawn on the maps (Table I). However, we discovered significant results concerning sketch map scale, average number of elements, element interdependency and types of landmarks identified.

TABLE I. AVERAGE NUMBER OF ELEMENTS

	Gender		Prior Experience	
(total maps)	M (61)	F (48)	Yes (26)	None (83)
All Elements	15.97	14.04	14.27	15.39
Paths	5.38	4.81	4.65	5.28
Edges	1.44	1.29	1.38	1.37
Districts	1.84	1.71	1.65	1.82
Nodes	1.77	1.58	1.5	1.75
Landmarks	5.54	4.65	5.08	5.17

A. Map Scales

Participants drew their sketch maps at a variety of scales ranging from 0.12 to 170 square miles, but the distribution was not even. After arranging the sketch maps in ascending order based on area, we analyzed the percent change between subjects' maps and noticed over a 100% change in area from 8 to 19 square miles, and as a result, we developed two catgories of maps: neighborhood level and citywide level (Fig. 1).

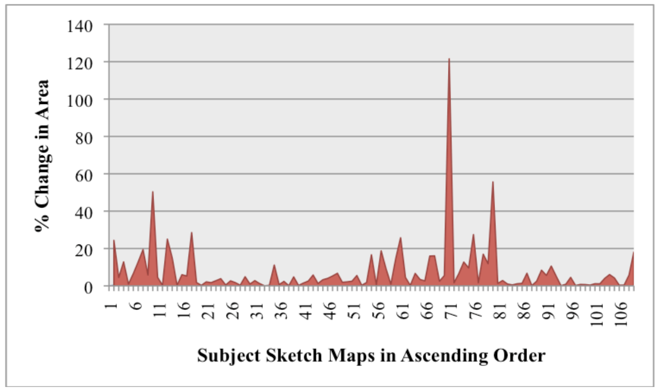


Figure 1. Percent change in area from subject to subject using 109 sketch maps arranged in ascending order. Given the over 100% percent change between subject 70 and 71, we divided the subject's maps into two groups: neighborhood level (70 total) and citywide (39 total).

A total of 70 neighborhood level sketch maps had an arithmetic mean of 2.02 square miles, with a standard deviation of 2.02. The 39 total citywide sketch maps had an arithmetic mean of 85.4 square miles and a standard deviation of 39.6. The standard deviations serve as indicators as to how closely the data cling to the sample mean. A small standard deviation indicates a set of data distributed tightly around the mean, whereas a large standard deviation indicates a wider range of areas across the sample maps.

We were surprised by the emergence of two distinct scales; nothing in the protocol suggested any bias toward the scale of the maps. We believe that this represents an inherent inclination on the part of the subject toward viewing the city as either a small comprehensible unit or a large extension in space.

B. Average Number of Elements

One of our central questions was how many elements would generally be understood as necessary for a competent sketch map of a large urban environment. Our evidence in Table II and III points to a relatively limited number of total elements: all sketch maps combined revealed an average of 15.12 (+/- 7.48); neighborhood level maps 14.2 (+/- 7.83); and citywide level maps 16.77 (+/- 6.58). We believe that the number of total elements is related to two factors: the need to provide clear representation and the limits of cognitive recall.

The amount and clarity of spatial information that can be presented is probably of particular concern to our subject group (design students), but it is likely also an issue to any user. We believe that this number of elements can be understood as each user's best guess of what would be legible to a third party.

The limits of cognitive recall also tend to simplify sketch maps. Cognitive maps always have less detailed information than geospatially constructed representations; they represent information that can be quickly recalled. While this limits the amount of information, it is good guide to the limits of efficient recall.

 ISBN: 978-1-61208-250-9

TABLE II. AVERAGE NUMBER OF ELEMENTS FOR ALL SKETCH MAPS

	All Sketch Maps (109 maps total)
All Elements	15.12 (+/- 7.48)
Paths	5.13 (+/- 4.20)
Edges	1.38 (+/- 0.65)
Districts	1.78 (+/- 1.48)
Nodes	1.69 (+/-1.13)
Landmarks	5.15 (+/- 3.98)

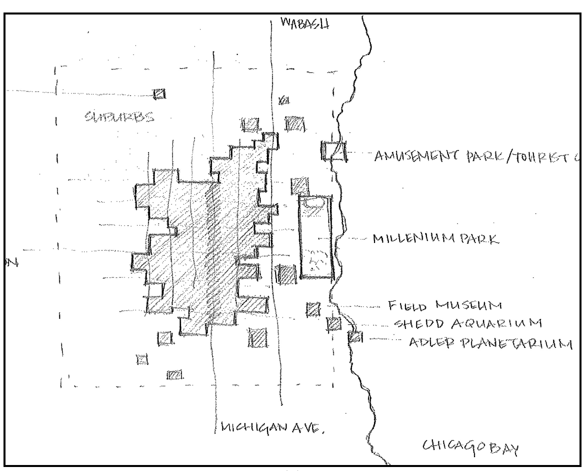
TABLE III. AVERAGE NUMBER OF ELEMENTS FOR NEIGHBORHOOD AND CITYWIDE SKETCH MAPS

	Neighborhood Level (70 maps total)	Citywide Level (39 maps total)
All Elements	14.2 (+/- 7.83)	16.77 (+/- 6.58)
Paths	5.17 (+/- 4.48)	5.05 (+/- 3.71)
Edges	1.3 (+/-0.67)	1.51 (+/- 0.6)
Districts	1.34 (+/- 1.27)	2.56 (+/- 1.52)
Nodes	1.41 (+/- 1.08)	2.17 (+/- 1.05)
Landmarks	4.97 (+/- 4.34)	5.46 (+/- 3.24)

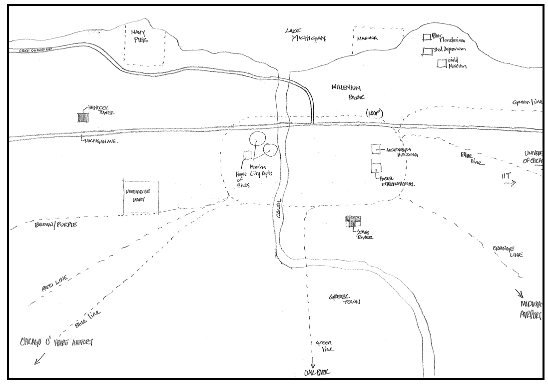
C. *Element Interrelationship*

We were interested in how the use of the Lynch's five elements would be managed by the subjects. We began the study with the hypothesis that some of the elements might be less important, or that one element category might quantitatively dominate the others. To our surprise, we found that all of the five elements were used in the sketch maps (although occasionally, a subject omitted an element type), and that Lynch's description seemed to fit the untutored sketch maps produced by the participants.

Second, using all 109 sketch maps, we found a significant inverse correlation between landmarks and districts. Subjects appeared to describe the city either in terms of landmarks or districts, but not generally as both. We are aware this finding may only be true for our subject population or for the city of Chicago, but it is suggestive that either landmarks or districts are sufficient to develop a cognitive map and that they both may be referencing similar spatial attributes. This would suggest that interface designers might use either one element or the other, but not need to duplicate them (Fig. 2a and 2b)



(a)



(b)

Figure 2. Two participant sketch maps showing the inverse relationship between districts (a) and landmarks (b).

Third, our qantitative analysis reveals a 91% increase in the amount of districts and 53.9% increase in the amount of nodes drawn in the citywide level maps when compared to the neighborhood level maps (Table IV). Districts are areas of a city, from several blocks to a large area, which share some kind of identifying characteristic and nodes are spots within a city which can be entered and at which different activities occur in multiple structures (e.g. parks, squares, shopping/entertainment centers, etc.). Based on these explanations, our finding suggests both districts and nodes serve as instruments for participants to reduce their cognitive workload when recalling the complexity of a citywide image.

TABLE IV. PERCENT CHANGE IN ELEMENTS

	Neighborhood	Citywide	% Change
All Elements	14.2	16.77	18.1%
Paths	5.17	5.05	2.3%
Edges	1.3	1.51	16.2%
Districts	1.34	2.56	**91%**
Nodes	1.41	2.17	**53.9%**
Landmarks	4.97	5.46	9.9%

D. *Semantic and Physical Landmarks*

We began the study with a hypothesis that the nature of landmarks has changed since Lynch's study. We have previously written about the emergence of a dispersed form of urbanism that relies less on architectural form and more on the flow of people and information [19, 20, 21].

Recalling that Lynch's definition of landmarks is based solely on formal attributes (e.g. a tower), we were specifically interested to see if subjects would identify landmarks based on their form or on their function. Our subject group would be if anything prejudiced as architecture students towards form; however, we found that overall 22% of landmarks are identified functionally (Table V).

Additionally, we found that the percentage of functionally defined landmarks was significantly higher for neighborhood level maps (27%) than for citywide (15%) (Table VI). We speculate that the finer grain of the neighborhood level maps explains the higher use of functional landmarks, but even at the citywide level, Lynch's definition of landmarks based on form neglects semantic landmarks which appear with significant frequency. We believe the amount of functionally significant landmarks identified in the maps confirms our original hypothesis about the changing nature of urbanism, even within a more traditional city such as Chicago.

TABLE V. RECALL OF PHYSICAL VERSUS FUNCTIONAL LANDMARKS

	Physical Presence	Functional Significance
Landmarks	77%	22%
Average	4 (+/- 3.06)	1.15 (+/- 1.9)

TABLE VI. PHYSICAL VERSUS FUNCTIONAL LANDMARKS

	Neighborhood Level		Citywide Level	
	Functional	Physical	Functional	Physical
Landmarks	27%	73%	15%	85%
Average	1.34	3.63	0.8	4.67

V. CONCLUSION AND DESIGN GUIDELINES

The use of cognitive maps in urban visualization impacts at least two types of interfaces. The most obvious impact is on what we might call “heads up” applications. These applications are for devices that must convey a maximum amount of information as quickly as possible. Mobile devices of all types, including but not limited to GPS devices, must not distract drivers or users from their immediate task. Cognitive maps can provide interface designers with a quantitative understanding of the most important elements and their relationship, helping to set limits on the numbers and type of elements.

By contrast, urban mapping and visualization may seem unlikely candidates because of the immense amount of data (layer upon layer of information about roads, buildings, flood zones and businesses) and the heterogeneity of the information. But here too, cognitive maps provide us with insights about the layering of information. Given a complex set of data, an interface designer will need to sift through this wealth of information to foreground the most important and provide a hierarchy of primary, secondary and tertiary information within a densely packed interface. Cognitive maps, because of the variety of elements, also provide one method to combine spatial and semantic information. Thus, for complex systems, cognitive maps provide us with guidance for “drilling down” into complex information.

Design guidelines for mapping interfaces:

- The number of elements used in interactive maps should be limited to approximately 15. Top level maps should guide user interaction.
- Landmarks and districts are redundant and a choice should be made to use one or the other.
- Maps are likely to be understood at the citywide level or the neighborhood level; good design will incorporate both scales into the interface.
- Designers should consider including both physical and functional landmarks in their maps.

REFERENCES

[1] K. Lynch, Image of the City. Cambridge: MIT Press, 1960.

[2] B. Tversky, “Cognitive maps, cognitive collages, and spatial mental models,” in Spatial information theory: A theoretical basis for GIS, A. U. Frank and I. Campari, Eds. Berlin: Springer-Verlag, 1993, pp. 14-24.

[3] D. Appleyard, Planning a Pluralist City: Conflicting Realities in Ciudad Guayana. Cambridge: MIT Press, 1976.

[4] D. Appleyard, “Styles and methods of structuring a city,” Environment and Behavior, vol. 2, no. 1, 1970, pp. 100-117.

[5] M. Agrawala and C. Stolte, “Rendering effective route maps: improving usability through generalization,” Proc. SIGGRAPH Symp. Computer graphics and interactive techniques, ACM Press, 2001, pp. 241–249. doi:10.1145/383259.383286.

[6] H. Ziegler and D. Keim. “Copernicus: Context-preserving engine for route navigation with interactive user-modifiable scaling,” Proc. Eurographics/ IEEE-VGTC Symp. Visualization, Blackwell Publishing, vol. 27, no. 3. 2008, pp. 927-934.

[7] J. Lee, J. Forlizzi, and S. E. Hudson, “Iterative design of move: A situationally appropriate vehicle navigation system,” International Journal of Human-Computer Studies, vol. 66, no. 3, Duluth, MN: Academic, 2008, pp. 198–215, doi:10.1016/j.ijhcs.2007.01.004.

[8] F. Grabler, M. Agrawala, R. Summer, and M. Pauly, “Automatic generation of tourist maps,” Proc. SIGGRAPH 2008 Transactions on Graphics, ACM Press, vol. 27, no. 3, Aug. 2008, Article 100, doi: 10.1145/1360612.1360699.

[9] G. Leshed, T. Velden, O. Rieger, B. Kot, and P. Sengers, “In-car gps navigation: engagement with and disengagement from the environment,” Proc. 26th annual SIGCHI conference on Human factors in computing systems (CHI ’08), ACM Press, April 2008, pp. 1675–1684, doi: 10.1145/1357054.1357316.

[10] S. Graf, W. Spiessl, A. Schmidt, A. Winter, and G. Rigoll, “In-car interaction using search-based user interfaces,” Proc. 26th annual SIGCHI conference on Human factors in computing systems (CHI ’08), ACM Press, April 2008, pp. 1685–1688, doi: 10.1145/1357054.1357317.

[11] I. M. Jonsson, H. Harris, and C. Nass, “How accurate must an in-car information system be?: Consequences of accurate and inaccurate information in cars,” Proc. 26th annual SIGCHI conference on Human factors in computing systems (CHI ’08), ACM Press, April 2008, pp. 1665-1674, doi: 10.1145/1357054.1357315.

[12] E. Tolman, “Cognitive Maps in Rats and Men,” Psychological Review, vol. 55, no. 4, 1948 pp. 189–208.

[13] R. Briggs, “Urban Cognitive Distance,” Image and Environment, R. M. Downs and D. Stea, Eds. Chicago: Aldine, 1973, pp. 361-388.

[14] A. Rapport, The Meaning of the Built Environment: A Non-Verbal Communication Approach, Tucson: University of Arizona Press, 1990.

[15] P. R. Gould, “On Mental Maps,” in Image and Environment, R. M. Downs and D. Stea, Eds. Chicago: Aldine, 1973, pp. 182-220.

[16] R. G. Golledge, “Learning about Urban Environments,” in Timing Space and Spacing Time, T. Carlstein, D. Parkes and N. Thrift, Eds. London: Edward Arnold, 1978 pp. 76–98.

[17] W. G. Evans, D. Marrero, and P. Butler, “Environmental learning and cognitive mapping,” in Environment and Behavior, vol. 13, no. 1, Jan. 1981, pp. 83–104, doi: 10.1177/0013916581131005.

[18] D. Hintzman, C. O’Dell, and D. Arndt, “Orientation in cognitive maps,” Cognitive Psychology, vol. 13, no. 2, April 1981, pp. 149–206, doi: 10.1016/0010-0285(81)90007-4.

[19] G. Wessel and E. Sauda, “Urban User Interface,” Proc. ACSA Interational Confernce, Barcelona, Spain, June 2012.

[20] G. Wessel, C. Ziemkiewicz, R. Chang, and E. Sauda, “GPS and Road Map Navigation: The Case for a Spatial Framework for Semantic Information,” Proc. International Advanced Visual Interface Conference, Rome, Italy, 2012, pp. 207-214.

User Support System for Designing Decisional Database

Fatma Abdelhédi, Gilles Zurfluh
University of Toulouse1 Capitole – IRIT *(UMR 5505* - France)
{Fatma.Abdelhedi, Gilles.Zurfluh}@irit.fr

***Abstract*— The design of a multidimensional schema is usually performed by a specialist (computer scientist). According to data-driven, requirement-driven or hybrid-driven approaches, he determines the facts and axis of analysis. Such an approach assumes that the decision maker expresses, more or less formally, analysis needs and communicate them to the computer scientist. We propose multidimensional schema designing by the decision maker himself following a hybrid-driven approach. Through a process of assistance successively viewing intermediate schemas from sources, the decision maker gradually built his multidimensional schema. As what determined the measures studied, the analysis dimensions and hierarchies within dimensions. A software tool named SelfStar based on this principle has been developed and validated with decision makers.**

Keywords-Multidimensional model; design process; decisional Data-base; decision-makers' requirements; data-source

I. INTRODUCTION

Decisional Data Base (DDB) allows decision maker to analyze data structured according to a multidimensional schema (star or constellation). Multidimensional schema design has motivated numerous works which can be derived from data-driven, requirement-driven or hybrid-driven approaches [1][2][3]. These approaches, to perform as well as they are, all involve a designer who is only able to design a multidimensional schema from decision makers' requirements or data source schema. The development of such a schema is a cumbersome process because it requires decision makers to articulate their needs to a designer and that it captures well the need to translate them to a adapted schema form.

Decision makers do not generally mastered formalisms and IT (Information Technology) tools but are experts in their field work. The SelfStar project fits into this fact and aims to define the data warehouse process dedicated to casual users (decision makers). It propose an approach and a software tool allowing to the decision-maker to design himself his DDBs incrementally, according to his requirements, and this without recourse to an expert designer. This paper focuses on the constellation schema [4] design.

This paper focuses on the incremental design of a constellation schema from the relational schema of a data source. This work is part SelfStar project developed by our team and aimed to decision-makers to design their data-warehouses.

The paper is structured as following. In Section 2, we present briefly several approaches to design a multidimensional schema. The justification of SelfStar system is provided in Section 3. In Section 4, the input and the output of the system are successively defined. Section 5 is devoted to a presentation of multidimensional schema process. Section 6 describes the architecture of a case tool allowing to experiment these mechanisms.

II. RELATED WORK

Many studies have been devoted to the multidimensional schema approach. These approaches can be classified into 3 categories: data-driven, requirement-driven, and hybrid-driven approaches. Data-driven approach [1][5] uses database source to generate a set of candidate multidimensional schemas and presents the drawback of not taking into account decision makers' requirements, consequently, going towards a possible failure of expectations of decision maker. Requirement-driven approach [2][6] takes into the needs of decision makers, but the sources are ignored at first. After that, multidimensional schema has to be mapped on the data sources. So, the designer can discover that the requirements do not correspond to sources.

Hybrid-driven approaches [3][7] combines and integrate the benefits of these two previous approaches. In fact, this approach designs on the one hand candidate schemas from the data (data-driven approach) and on the other hand multidimensional schema from decision makers' requirements (requirement-driven approach). A designer must map these two types of schemas to obtain a consistent multidimensional schema. Romero and Abelló [3] propose an automatic method Multidimensional Design by Examples (MDBE) following a hybrid-driven approach. To generate multidimensional schemas, this method takes as input, on one hand, the decision makers' requirements expressed as SQL queries, and on the other hand, the relational data source.

Source analysis is provided by SQL queries and knowledge of relational source schema. Therefore, the multidimensional schema design requires an expert (a computer scientist) to formulate SQL [15] query and analyze data source. Pinet and Schneider [8] propose to generate a multidimensional schema from a conceptual schema using UML notations. This approach represents source classes with a directed acyclic graph. The user chooses a node from this graph to design a fact. All connected nodes to this chosen fact represent the potential dimensions of this fact. However, in our opinion, this

 ISBN: 978-1-61208-250-9

representation of multidimensional schema is complex for the decision maker because of the number of generated nodes to represent the dimension hierarchies. However, interactions with the decision maker choices are limited to the facts.

To our knowledge, few works [7] try to allow users to participate in the process of multidimensional schema.

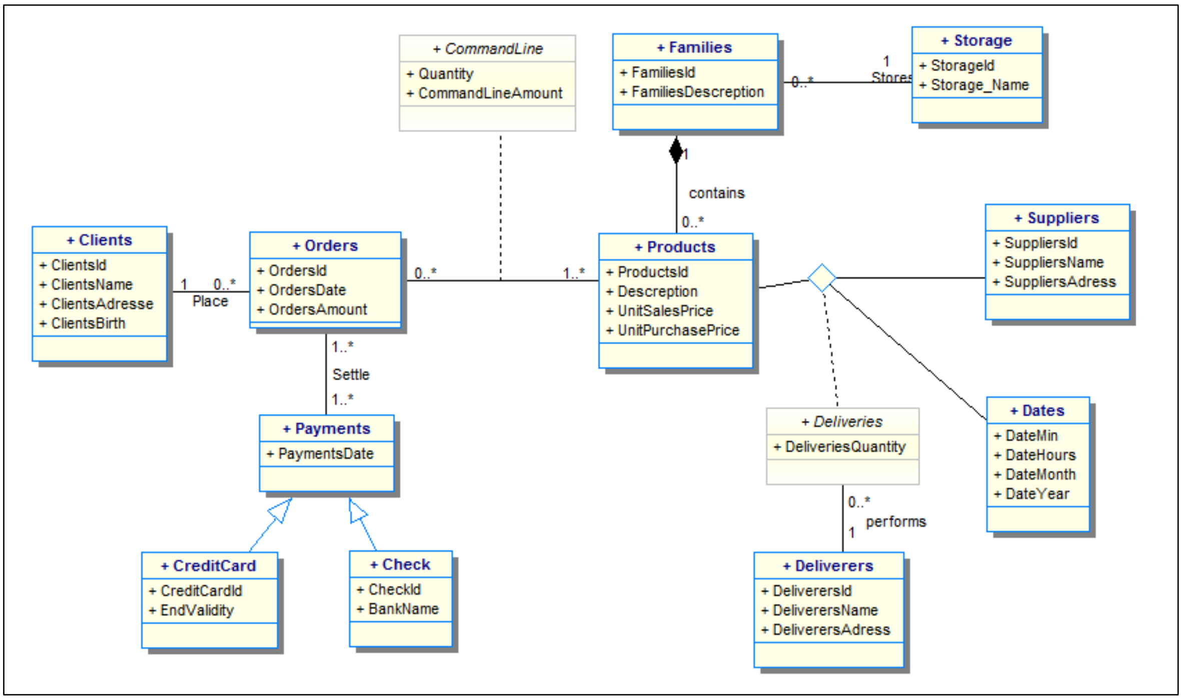


Figure 1. UML Class Diagram of products sales ans stock management

III. JUSTIFICATION OF SELFSTAR PROJECT

In the current approach, the decision makers should ask the specialists of the data management (administrators, computer scientist) each time they wish to get a new decisional database or to evolve the deprecated multidimensional schemas. This is related primarily to the complexity of:

- principles of designing of a decisional database
- ETL Processes ensuring the periodic load of the decisional database from there [9][10].

SelfStar project aims to propose a full approach and a software environment to allow to the decision maker to design a constellation schema. The design process of the schema is based on a hybrid approach: it starts from the database schema (class diagram CD) and the decision makers' requirements. This process is incremental: the decision maker integrates progressively the analysis needs in 3 successive intermediate schemas.

The advantage of such process is twofold:

- The decision maker becomes independent to express his decisional requirements,
- The process automatically controls the correspondence between analysis requirements and sources.

However, the decision maker, even if he knows his needs, is obviously confronted with a double complexity:

- Data source organization (Relational, Entity-Association or UML diagram),
- Process of the star or constellation schema design (fact(s), dimensions, hierarchy).

SelfStar project aims to propose formalized mechanisms to palliate this complexity.

IV. INPUT AND OUTPUT OF SELFSTAR

SelfStar system allows constellation schemas design from a data source and the decision-maker's requirements. In order to illustrate the input/output of the design process, we present successively an example of source schema, an expression of analysis needs and finally the decisional base schema resulting from the process.

A. Data source: Input of SelfStar

Data source is described by UML language. The choice of this formalism is justified by the semantic richness of the data model [2] and also by the correspondence between this language with the models entity-association and relational. Figure 2 gives the example of a classes diagram (CD) describing a management of the sales.

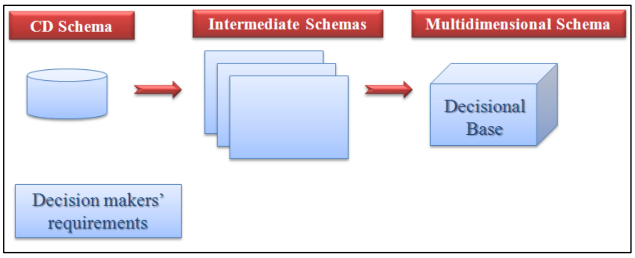


Figure 2. Our design process that allows a decision-maker to build a data-mart schema

The CD not being easily exploitable in the state, it is the subject thus of a reprocessing. This principle was already proposed in [11] for the Entity-Association model and in [7] for the UML CD. According to this principle, the conceptual schema is transformed into a simplified schema containing only object classes and binary links of type 1..N between these classes. The transformation of a DCL into an exploitable schema is carried out as follows:

- An object's classes or association's classes of the source becomes a class in the exploitable schema,
- A binary association links type 1..N is deferred in the state,
- A binary association links type M..N is transformed into a class bound by 2 links 1..N,
- A link of aggregation or composition is processed as an association links (these type of links are not meaningful in the multidimensional schema),
- A link of heritage disappears; the subclass of the source becomes a class and it will be on the same level as the super-class by inheriting its attributes and links (it preserves the semantics of the data).

The exploitable schema obtained is an oriented acyclic graph.

B. *Requirements: Input of SelfStar*

When a decision maker wants to analyze the data of a source, it can express his requirements:

- Either in the form of SQL request [12] ; but this proves to be difficult for decision maker except IT, especially when these requests make use of clauses Group By and Having;
- Or in the dashboards forms corresponding to the expected analysis results.

In SelfStar, the decision maker integrates his requirements by himself through a graphic interface and this without formalizing them beforehand.

C. *Data-mart: output of SelfStar*

From a data source and analysis needs, SelfStar system visualizes a decisional database on which the decision maker can apply its queries. The schema of Figure3 describe a decisional database allowing the decision maker to analyze the number of Orders (Orders_Nb measure) and the delivered quantity accumulated (Orders Amount measure) according to Products and Dates dimensions. Each dimension is associated to a set of organized parameters into hierarchies.

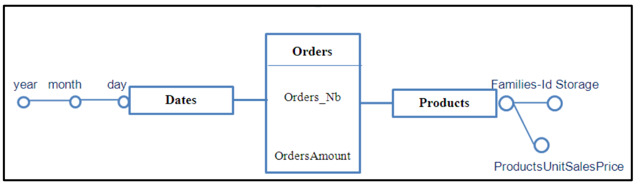


Figure 3. Data-mart schema

V. PERSONNALIZATION PROCESS

A. *Multidimensional schema design*

The proposed process in SelfStar (Figure 3) has a starting point the data source schema (CD) and the decision-makers' requirements (informal). It contains four successive steps in which the decision maker will interact with the system to integrate progressively his requirements. Each step produces a new schema more comprehensive than the previous one. The 4th schema corresponds to that of the decisional database, i.e. the desired result (Figure 3). The decisional schema design is performed incrementally.

First step consists to display a set of candidates facts in the intermediate schema number 1 (noted IS1). InIS1, the decision maker chooses the fact that he wants to analyze with the measures and their aggregations functions.

Second step generates automatically the intermediate schema number 2 (IS2); it proposes all possible associated dimensions with the selected fact. InIS2, the decision maker will be able to indicate dimensions according to which they wish to analyze the fact.

Third step generates an intermediate schema number3 (IS3) containing the constellation schema (fact(s) + dimensions) with all the possible hierarchies. In IS3, the decision maker will choose each responding hierarchies to their requirements.

Fourth step produces the multidimensional schema data-mart. At this level SelfStar will record the personalization metadata which later will allow the decision maker to elaborate the newest multidimensional

ISBN: 978-1-61208-250-9

schemas. All algorithms related to the different phases of the process are presented in [13].

B. Personnalization meta-data

Industrial experience of our team [14] showed that (1) a data source (sales or production BD) contains frequently from 30 to 60 objects classes and many links and (2) the analyzes carried by the same decision maker are usually very similar in terms of facts and dimensions.

According to (1), we consider that the decision maker isn't able to choose the fact directly from the conceptual source schema because this schema is considered complex for any person (except computer scientist). We have decided to show for the decision maker a simplified presentation of the source (noted IS1) that is extracted automatically from the source schema.

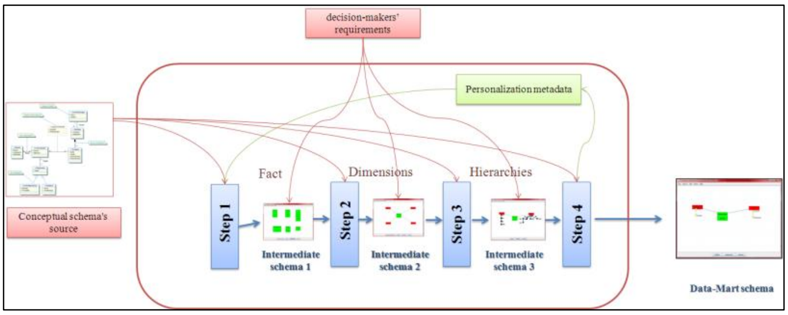


Figure 4. The transformation from data-source and decision makers' requirements to data-mart schema

The second point noted (2), led us to define a personalization mechanism on the IS1; this mechanism can extract, from the CD of source, the most significant classes for helping the decision maker.

The simplification of a source is based on the extraction of classes which must be analyzed by the decision maker (candidates facts). This extraction method is based on the personalization technical developed in our team and which inspired from the work of [12]. SelfStar record the personalization metadata whenever a decision maker develops a decisional schema. It is based on the scheduling of the candidate facts when the same decision maker will design a new multidimensional schema. Figure5 presents the algorithm that producing the personalization metadata.

```
Algorithm : PersoWeight
Input : ES, CW, MS, U        -- Exploitable Schema, Classes Weight,
Multidimensional Schema and User
Output: CW                    --new weight values
begin
for i←1 to p do              -- from each fact from MS
  s ← source(Fi)
Ps← Ps+ Coef   -- increases the class source weight
for x in Dim(Fi) do
   s ← source (x)
   Ps ← Ps +2
end for
end for
```

Figure 5. Generation personalization metadata

VI. Case Tool

A. Experimentation

SelfStar project was implanted in order to experiment all the proposed mechanisms. Figure 6 presents the software architecture. JAVA was used to develop this software. We used JAXP (Java API for XML Processing) that is a set of API including the SAX, DOM, XSLT (eXtensible Stylesheet Language Transformations) and XPath. The set of schemas used by SelfStar are with XML. We use JAXP tools to transform them. To view all XML schemas (intermediate schemas), we have developed a visualization module. It allows also the interaction between the system and the decision maker via the interface. This module uses the JGraph library.

Soon activation, the software demands to the decision maker identifier of relational database (source). CD source is transformed into an exploitable schema through the XSLT API (XSLT generator). The DOM generator creates then intermediate schema 1 (IS1) through by analyzing exploitable schema and using basic metadata (calculated weight). IS1 is displayed and the decision maker can choose directly on the screen fact(s) to analyze. This process is repeated to integrate choice of dimensions and attributes in the following intermediate schemas (IS2 and IS3).

Finally, the multidimensional schema will be displayed and personalization metadata are generated. They will subsequently be used to guide the decision maker in developing new multidimensional schemas. It should be noted that during the first analysis of the source, only the basic metadata are available. With each new

 ISBN: 978-1-61208-250-9

analysis of this source, the metadata is enriched and will better define the profile of the decision maker.

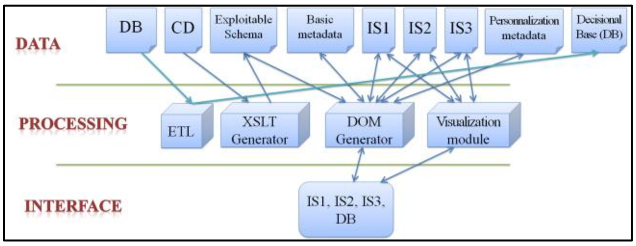


Figure 6. Software architecture of SelfStar

This paper does not deal ETL process. But, SelfStar has a module for loading data warehouse from sources. This module is automatically generated by SelfStar after the development of multidimensional schema.

B. Validation

To validate our approach and our tool, we used marks management application at University of Toulouse. The database source is composed of extracted data from the management education package (APOGEE [16] available in most french universities). We chose three decision makers of staff responsible for managing marks and awarding degrees.

The source is a relational database containing training descriptions, students and results. Figure 7 shows an extract from the schema describing the database source.

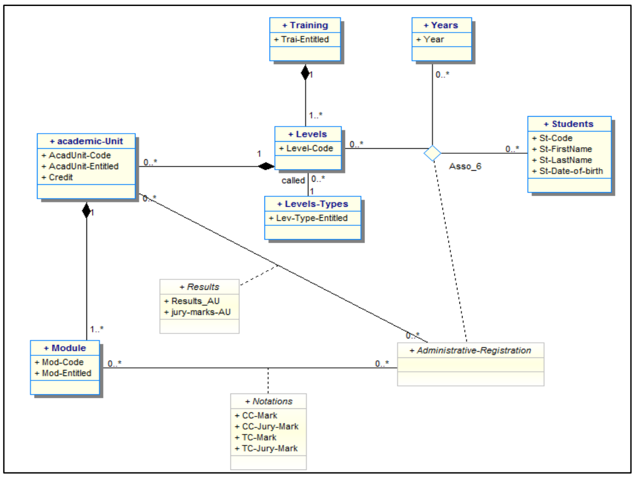


Figure 7. *CD extract of Database source*

We asked three employees of the University, having similar activities but independent create their own decisional databases using the database sources for analysis. It is noted that these decision makers use Business Object [17], occasionally.

(1) An official school: analyzes the students' absenteeism in the exams; observes the dates of issuance of marks of exams, etc.

(2) A training manager: defines the rate of exam success by students, analyzes the evolution of result per scholar year, etc.

(3) An academic manager of an academic year: examines student grades per subject per teacher (average, standard deviation, etc.), analyzes the results per subject, etc

After a brief training regarding the using of SelfStar, these three decision-makers showed certain affluence to quickly design new data-marts and manipulate them with Business-Object. The scheduling of the candidate facts was realized through the mechanism of personalization.

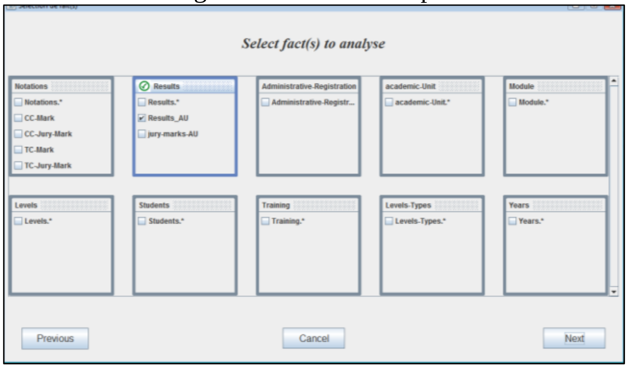


Figure 8. *Intermediate schema n°1*

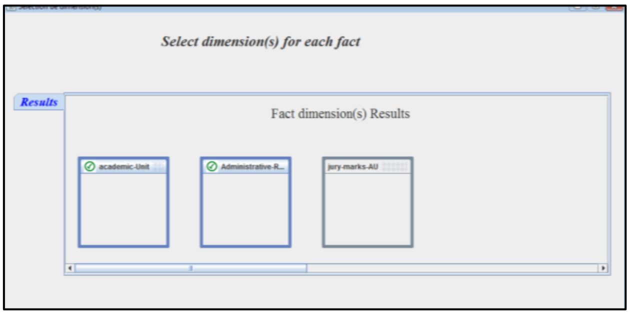


Figure 9. *Intermediate schema n°2*

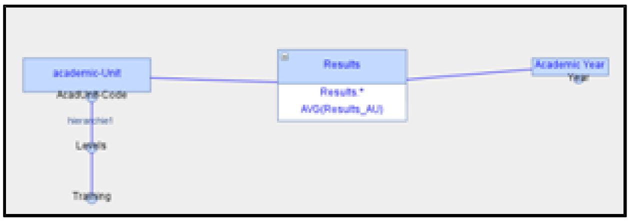


Figure 10. *Multidimensional schema*

VII. CONCLUSION AND FUTURE WORK

This paper provided an approach to design a multidimensional schema from the data sources schema to be analyzed; the decision makers' requirements are integrated progressively throughout the process. This approach is original in the measurement that it allows for the decision maker to design progressively his multidimensional schema without resorting to a database specialist (computer scientist). It is distinct clearly from the others data-driven, requirement-driven or hybrid-driven approaches in which the user is not directly involved.

References

[1] F. Ravat and O. Teste, « Personalization and OLAP databases », New Trends in Data Warehousing and Data Analysis, vol. 3 , pp. 1–22, 2009.

[2] N. Prat, J. Akoka, and I. Comyn-Wattiau, « A UML-based data warehouse design method », Decision Support Systems, vol. 42, no 3, pp. 1449–1473, 2006.

[3] O. Romero and A. Abelló, « Automatic validation of requirements to support multidimensional design », Data & Knowledge Engineering, vol. 69, no 9, pp. 917–942, December, 2010.

[4] R. Kimball, « The data warehouse toolkit : Practical Techniques for Building Dimensional Data Warehouses », John Wiley and Sons, ISBN : 0-471-15337-0, 1996.

[5] D. L. Moody and M. A. R. Kortink, « From enterprise models to dimensional models: a methodology for data warehouse and data mart design », DMDW'00, Sweden, vol. 5, 2000.

[6] J. Trujillo, S. Lujan-Mora, and I. Y. Song, « Applying UML and XML for designing and interchanging information for data warehouses and OLAP applications », Journal of Database Management (JDM), vol. 15, no 1, pp. 41–72, 2004.

[7] P. Giorgini, S. Rizzi, and M. Garzetti, « Goal-oriented requirement analysis for data warehouse design », in Proceedings of the 8th ACM international workshop on Data warehousing and OLAP, 2005, pp. 47–56.

[8] F. Pinet and M. Schneider, « A unified object constraint model for designing and implementing multidimensional systems », Journal on Data Semantics XIII, pp. 37–71, 2009.

[9] P. Vassiliadis, « A survey of Extract–transform–Load technology », International Journal of Data Warehousing and Mining (IJDWM), vol. 5, no 3, pp. 1–27, 2009.

[10] F. Atigui, F. Ravat, O. Teste, and G. Zurfluh, « Using OCL for Automatically Producing Multidimensional Models and ETLProcesses », Data Warehousing and Knowledge Discovery, pp. 42–53, September 2012.

[11] I. Y. Song, R. Khare, Y. An, S. Lee, S. P. Kim, J. Kim, and Y. S. Moon, « SAMSTAR: An automatic tool for generating star schemas from an entity-relationship diagram », Conceptual Modeling-ER 2008, pp. 522–523, 2008.

[12] F. Ravat, O. Teste, R. Tournier, and G. Zurfluh, « Graphical querying of multidimensional databases », in Advances in Databases and Information Systems, pp. 298–313, 2007.

[13] F. Abdelhédi, G. Pujolle, O. Teste, and G. Zurfluh, « Computer-Aided Data-Mart Design », ICEIS, pp. 239-246, June - 2011.

[14] E. Annoni, F. Ravat, O. Teste, and G. Zurfluh, « Towards multidimensional requirement design », Data Warehousing and Knowledge Discovery, pp. 75–84, 2006.

[15] « Accueil oracle ». [Online]. Available: http://docs.oracle.com/cd/B19306_01/server.102/b14200/toc.htm. [Accessed: 12/25/2012].

[16] « Accueil Amue - Amue ». [Online]. Available: http://www.amue.fr/. [Accessed: 12/25/2012].

[17] « SAP France - SAP BusinessObjects - Solutions de Business Intelligence (BI) et de gestion de la performance ». [Online]. Available: http://www.sap.com/france/solutions/sapbusinessobjects/index.epx. [Accessed: 12/25/2012].

Information Needs of Chinese Mobile Internet Users

A User Study

Yanxia Yang
TWYWI Consulting,
San Jose, CA, USA
marsha.yang@gmail.com

Grace Deng
TWYWI Consulting
San Jose, CA, USA
graceyd2006@gmail.com

***Abstract*—This paper investigates the information needs of Chinese mobile Internet users in their fast-paced environment. The mobile Internet users grown up in such an environment would have different interests, ways of using, and information needs. In order to obtain a better understanding on Chinese mobile Internet user behavior, a web survey was conducted in Xi'an, China, followed by the previous ethnographical study. With the pervasiveness of smart-phones, people from different age groups joined the mobile Internet market, and the results of this study illustrated the various user behaviors on the mobile Internet across different age groups of people. The study revealed that in the mobile Internet era, Chinese users are more concerned with information relating to the scope of safety and love/belonging needs, advancing towards the next level in Maslow's Hierarchy of Needs.**

Keywords-Internet; mobile Internet; interests; user behavior; web survey; mobile applications.

I. INTRODUCTION

Mobile applications have emerged as one of the most popular sector of the mobile industry chain. As shown by the newest statistics published by EnfoDesk databases, in the third quarter of 2011, the number of mobile Internet users in China has reached 396 million users, increasing at a rate of 4.4% [1].

The following graph shows the number of new iOS and Android device activations in China at the beginning of year 2012, an astounding growth which has surpassed that of the United States to become number one around the globe.

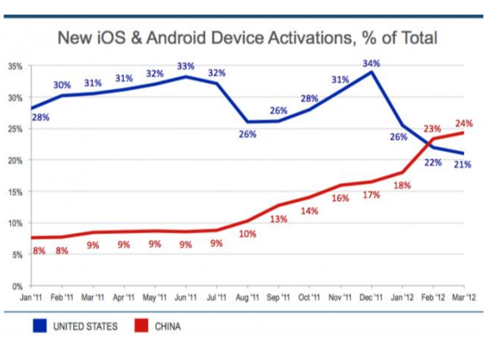


Figure 1. iOS and Android Device Activations in China and USA [2].

Based on Flurry's data in Figure 1, from Q1 2011 to Q1 2012, China led in the mobile app session growth as well, increasing 1,126% year-over-year. In such a fast-paced environment, users have various needs for the mobile applications. In order to better design mobile applications, we need to understand the information needs of mobile users. Thus, we designed a questionnaire and conducted a web survey in Xi'an, China.

The next section of this paper will discuss the methodology prescribed in questionnaire design and survey procedure. The third section will discuss findings and results of the web survey. The final section will analyze and summarize our findings.

II. METHODOLOGY

The research plan consists of ethnographic interviewing and web surveys. Ethnographic interviews were first conducted in Xi'an [3]. Afterwards, a web survey on customer concerned issues was administered as the second step of the research. This paper focuses on the second step.

According to the best practice suggested by Kuniavsky [4], the study plan was drafted to collect data to validate research hypothesis. According to the previous ethnographical study, Chinese Internet users focus more on two levels of Maslow's Hierarchy of Needs [3], safety and love/belonging, when surfing the Internet. We hypothesize that this same user behavior and focus on safety and love/belonging needs is demonstrated when these users utilize the mobile Internet. Out of these two levels, we wish to further investigate which subcategories are more important to the users. What specific information within these categories do they care about the most? Based on the previous study, we know that people have interests on trust, safety, education, job, psychology, communication, and health, etc., so this paper will study more on these interests.

A. Questionnaire Design

We designed the questionnaire based on the ethnographic results: Internet user behavior correlated to the user age group and other user demographic information. Since we assume that the users' information needs while using the mobile Internet are similar to their using the Internet, therefore we designed questionnaire focusing on the three basic needs in Maslow's Hierarchy of Needs: physiological, safety, and love/belonging.

The questionnaire includes both a first level factor and a second level factor. Each second level factor in turn encompasses multiple different related questions, and each question was designed for a 5-Lirket Scale answers from most agree to most not-agree. There are also some multiple choices or open ended questions that allowed survey participants to provide more inputs.

The first level factors are seven broad categories, such as Trust; Communication; Safety; Education; Jobs and Career; Psychology; and Health. The first level factors are further specified and divided by second level factors. For example, under the first level factor “Trust”, there are several second level factors such as “Consumer Recommendation”, “Shopping and Merchandise”, “Trust in Society”, “Location-Based Information Acquirement”, “Traffic”, etc. The following describes a few examples of the details of the questionnaire.

Example1:
First level factor - Trust
Second level factor - Consumer Recommendation
Related Questions:

1) When I purchase expensive electronic products, I usually research online beforehand.
   1. Most Not Agree, 2. Not Agree, 3. Neutral; 4. Agree, 5. Most Agree
2) What information do you look for? Choose from the following options or add your input.
   __quality of product, __manufacturer reputation, __ market price, __consumer reviews, __after-sales service, or other______?

.

Second level factor – Shopping and Merchandise
Related Questions

1) I only go shopping at a fixed number of stores.
   1. Most Not Agree, 2. Not Agree, 3. Neutral, 4. Agree, 5. Most Agree

2) I wish to receive information about sales at local stores.
   1. Most Not Agree, 2. Not Agree, 3. Neutral, 4. Agree, 5. Most Agree

3) I wish to receive electronic coupons and vouchers instead of paper ones.
   1. Most Not Agree, 2. Not Agree, 3. Neutral, 4. Agree, 5. Most Agree

4) I wish to save time by checking for information regarding products I buy often on my smart-phone.
   1. Most Not Agree, 2. Not Agree, 3. Neutral, 4. Agree, 5. Most Agree

B. Questionnaire Validity

The web survey was conducted using a formally designed questionnaire and surveyed a total of 200 mobile Internet users in China. In accordance to a completely randomized design in order to yield the most accurate results, the questionnaire was sent to a polling agency, where it was reviewed and proofread. Afterwards, the questionnaire was sent to a variety of different websites over the Internet in order to sample a more diverse and random population. When prefixed percentage ratios were satisfied, the poll was closed and data analysis was conducted on the results.

C. Predictive Validity

Prediction validity refers to the effectiveness of the outcome of a survey to predict the results of future surveys or experiments. As a form of criterion validity, the validity of the results is established when compared against known criteria. The main purpose of this survey is to target the most common problems current users face when using the mobile Internet. We also hope to investigate some of the most common issues these users are concerned about as well.

When information needs based on geographical location are taken into account, careful analysis of the survey results reflects the information needs of current users in different situations and holds predictive value on a long-term basis.

D. Construct Validity

Construct validity refers to whether the questionnaire is comprehensive and takes into consideration all facets and elements of the questions. It also refers to the extent to which what was to be measured was actually measured, and whether the results of the survey can account for the survey’s accuracy and validity. The formal questions in this survey are derived from the previous ethnographical study and the analysis of the users’ most commonly faced problems.

E. Content Validity

Content validity refers to the effectiveness of translating factors surveyed into specific and relevant questions, and the level by which these questions through content can accurately and comprehensively reflect a surveyed factor. It additionally refers to the verification of whether the survey method measures what is expected, as well as whether such a measure accounts for all aspects of a social construct.

These questions in the survey have been improvised and refined by comparing the content validity between a preliminary survey and that of a formal survey. Only conducting questions that involve measuring the satisfactory level of users when they encounter problems cannot completely reveal the information needs that users are interested in or are concerned about.

Hence, when designing this questionnaire, we not only took into account issues that users may confront, but also verified the specific types of information these users want to see. This particular verification process is administered by allowing surveyed users to select types of information they are interested in from a list of options and following up with analysis of the most popular choices indicated by the subjects.

III. Findings and Results

A. *Demographical Information*

1) Age group

Out of the total 200 web survey participants, the chart below illustrates the age group distribution for people who were born after the 1960's, 70's, 80's and 90's:

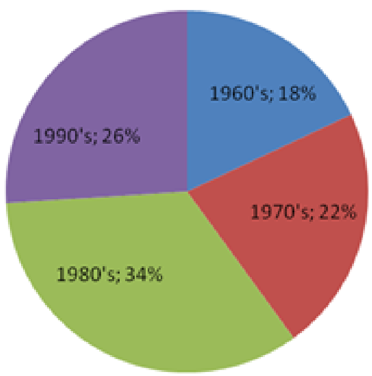


Figure 2. Age Group Distribution.

2) Sex

Out of the total web survey participants, 40.2% are female and 59.8% are male.

3) Marital status

Out of the total web survey participants, 15% were married with kids, 15% were married without kids; 30% were single and in a relationship; 40% were single and not in a relationship. The following chart shows this breakdown. The participants' marital status could influence their regularly-sought information. For example, if they are married with kids, their information seeking is more related to the family than if they were not married with kids, especially on issues such as childcare and children's education.

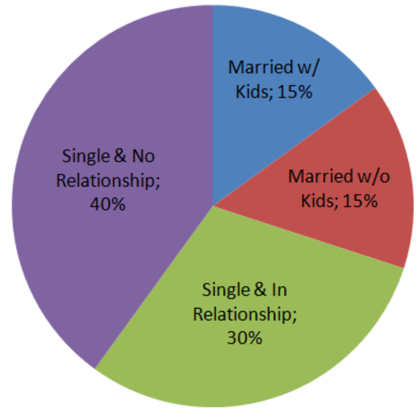


Figure 3. Marital Status Distribution.

4) Education background

Out of the total web survey participants, the education level breakdown is presented by the following chart:

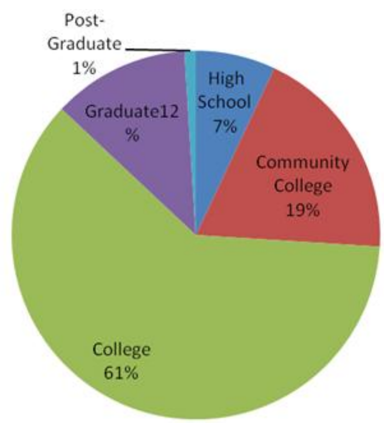


Figure 4. Education Level Distribution.

B. *Information Needs Related to the Gender*

By observing Figure 5, we find that a significant amount of women are more concerned with the job search category, compared to that of men because job security in China for women is lower than men due to the culture reason. Women are particularly interested in finding a job nearby. In another relevant category, career plan, it can also be concluded that women demonstrate more interest, although there is less of a statistical difference.

Another observation is the significant difference in interest between men and women on the education of children. On this topic, there is a greater difference than that of any other category. Men tend to like the information pushed to their mobile devices regarding how to educate their children, while women like to chat with their social

 ISBN: 978-1-61208-250-9

groups in order to get the information for children and education related topics.

Some of the common topics that both men and women are equally concerned about are those regarding food and nutrition safety, health and lifestyle, local events, and career planning.

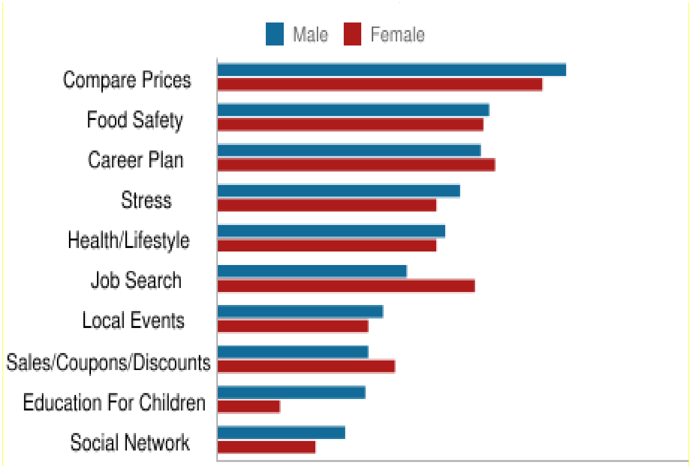


Figure 5. Interest in Different Information Categories Across Genders.

C. *Information Needs Related to the Age Groups*

The information needs broken down by age groups can be seen in the Table 1. The average Likert scale score for each type of questions were calculated according to the age groups.

TABLE I. INFORMATION NEEDS RELATED TO AGE GROUPS

Information Needs	Age Groups (born after …)			
	60s	70s	80s	90s
Compare Prices	4	4.45	4.59	4.56
Food Safety	4.4	4.64	4.27	4.22
Career Plan	3.6	4.36	4.3	4.78
Stress	4.4	4.36	4.13	4.44
Health/Life Style	3.8	4.27	4.13	4.56
Job Search	3.4	4	4.19	4.11
Local Events	4	3.82	3.95	3.89
Sales/ Discounts	3.4	3.55	4	4.22
Education for Children	4	4.55	3.65	4.11
Social Network	4.2	3.82	3.73	4.11

The above data can be illustrated in Figure 6. From this figure, it reveals that mobile Internet users born after the 60s are notably less concerned about sales/discounts, career planning and job search, compared to all other age groups, which indicated that this group has more financial freedom than other groups.

On the other hand, people born after the 90s are the most attentive to their career. In addition, this age group also display great attention to information related to the sales/coupons/discounts and health/lifestyle category, as compared to the three other age groups. A consistent trend shown by Figure 6 is that the younger age groups tend to be more concerned with physiological needs than safety needs, and pay more attention towards information that relate to livelihood (finding jobs and saving money).

For the people born after 70s, their main concerns focus on the food safety and education for children. Their needs are more related to safety and love/belonging as they have established social status. They like to check nearby restaurant reviews through their mobile devices before dining, and quickly learn from their mobile devices on others' or experts' opinions on how to educate their children.

A topic that evokes interest across all four age groups is that of news about local events and information regarding the local environment. They like to check local information through mobile devices on their way to commute in the public transportation, or in their spare time, because they think only local (specific to one city) information is most relevant to their daily lives.

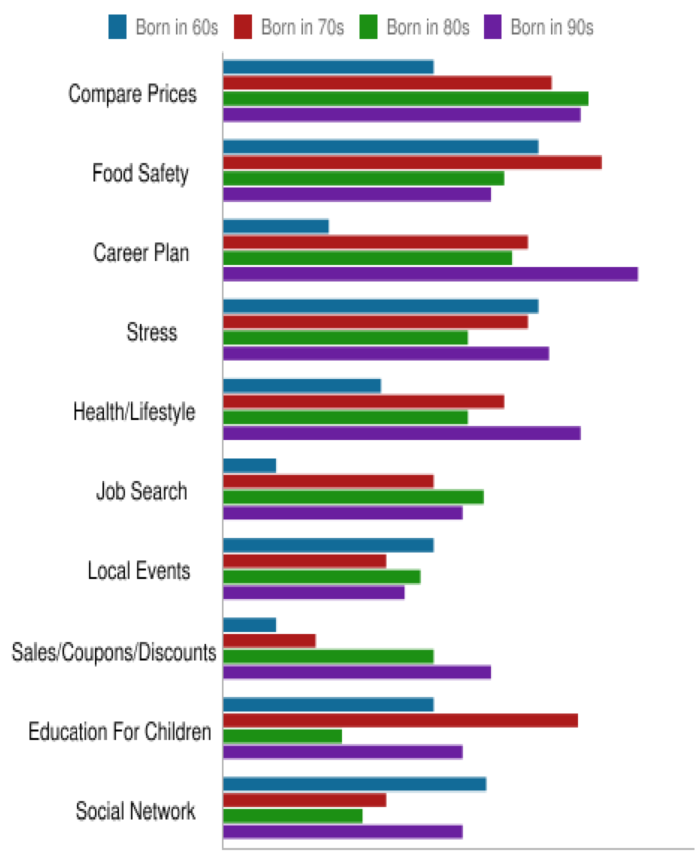


Figure 6. Interest in Different Information Categories Across Different Age Groups.

D. Local Information Needs-Checked with Mobile Devices Frequently

Figure 7 shows the top categories of information that surveyed mobile Internet users indicated what they wanted to know regarding their local community. As the chart illustrates, information related to saving money, such as discounts and sales, is clearly a high priority.

Users also like to utilize their mobile devices for quick access to daily local news and events happening in the neighborhood. They believe that the things happening around them are more meaningful to their lives.

Finally, many users also express equal interest in issues regarding livelihood, such as job openings and the price of gas and electricity. News regarding education also appears to be of considerate interest.

Figure 7. Local Information Needs.

IV. CONCLUSION

It can be concluded that mobile Internet users in China demonstrate information needs that are mostly on physiological, safety needs and love/belonging, and leaning towards the needs of safety and love/belongs. The mobile Internet users' information needs are the same as those on the Internet, and thus, also follow the egg theory by Yang [3]. By looking overall at Figures 5 and 6, the top categories of information that all mobile Internet users are concerned about as a whole are related to price comparison, food and nutrition safety, and career planning. Figure 7 also reinforces this observation. Categories that fall under physiological needs are that of price comparison. The mobile Internet introduced a quick access to people for online shopping and browsing. Categories that fall under safety needs are that of career planning, job search and self improvement, food safety, and health and lifestyle, which the users are more concerned with. Categories that full under the love/belonging are that of communication with friends and family, and child education, which have drastic use in the mobile Internet era.

Within the two categories of needs that mobile Internet users most focused on, safety and love/belong, which specific information do they care about the most? Analysis of Figures 5 and 6 reveals that the two most popular topics surveyed mobile Internet users were interested in are food safety, and career planning which all fall under the safety level of needs. The communal interest on information regarding the safety of food may be a result of the numerous reports of food poisoning or false products. Users may wish to go online through their mobile devices and check references or other customers' reviews of nearby restaurants' reputation and history before choosing those restaurants.

Here we elaborate more on how people's need relate to the level of love/belonging. The users communicate with their friends and family through cell phone calls, instant messaging, text, or wechat, which already has more than 200 million users, either to talk, share popular news/topics, or send personal pictures. Since a mobile phone's basic function is to communicate, the drastic increase in sales for mobile phones in China is due to a greater demand for faster, more accessible communication by the people. Many of these new mobile phone users are from the large amount of immigration of workers into the city from the countryside. In addition, college graduates often find jobs in bigger, well-developed cities than their home towns. This distance from their families contributes to these users' dependence on mobile phones to quickly connect with their families. The mobility within the cities also cause people to rely more on mobile phones than landlines because it is harder to keep a permanent landline number. Some people even have two or three phone numbers to reduce the cost of communication.

Finally, the demographical difference may cause different user behavior for the mobile Internet use. For example, as a result of the developing global economy, many job opportunities and career options have surfaced; therefore, many people are eager to acquire career information and counseling. This trend is especially prevalent in the female subset of surveyed subjects. This may be because women tend to want to find a job close to home in order to avoid traffic and to easily take care of their families. From a different perspective, the rankings given by people born in the 1960s show they are less concerned compared to the younger age groups regarding career planning and job search. As an overall trend, some of the older age groups tend to focus more attention towards information that fall under the safety category.

Human needs are a powerful resource of explanation of human behavior and social interactions [5, 6]. In the mobile Internet era, human needs have overall moved towards a higher level in the Maslow Hierarchy of Need [3]. People

 ISBN: 978-1-61208-250-9

are more concerned with things relating to the scope of safety and love/belonging. Between these two categories, different age groups of mobile Internet users have different methods of using the mobile Internet, as well as distinctive focused interests and objectives. There are several possible reasons that may contribute the variance, such as society development, influence of popular applications in the market during different times, education levels, amount of time spent on the Internet and mobile Internet, and many others.

The study of user behavior on the mobile Internet also reveals the extent by which mobile Internet is fusing into users' life. The closer mobile Internet is to its users' daily life, the more popular mobile applications will be. The next study will focus more on how users interact with the different types of mobile devices (phone/pad) on iOS, Andorid, and Win Mobile.

ACKNOWLEDGEMENT

Special thanks to the students of Xi'an Jiaotong University for their help to recruit participants.

REFERENCES

[1] EnfoDesk Databases, [Mobile Internet Users in China], Retrieved from http://www.enfodesk.com, October 2012.

[2] Flurry Analytics, [Graphic Illustration of iOS and Android Device Activations in China and USA]. Retrieved from http://www.flurry.com/flurry-analytics.html, June 2012.

[3] Y. Yang, H. Li, and G. Deng, "Case study: Internet and Mobile Internet User Behavior in China," Proc. HCII, Orlando, Florida, USA, 2011.

[4] M. Kuniavsky, "Observing the User Experience-A Practitioner's Guide To User Research," San Francisco, CA: Morgan Kaufmann Publisher, 2003.

[5] R.A. Coate and J.A. Rosati, "Preface," in The Power of Human Needs in World Society, ed. Roger A.Coate and Jerel A. Rosati, vol. IX, Boulder, CO: Lynne Rienner Publishers, 1988.

[6] A. H. Maslow, "A Theory of Human Motivation," in Psychological Review, vol. 50, pp. 370-96, 1943.

Emotion Recognition using Autonomic Nervous System Responses

Emotion Recognition

Byoung-Jun Park, Eun-Hye Jang, Sang-Hyeob Kim, Chul Huh, Myoung-Ae Chung
IT Convergence Technology Research Laboratory
Electronics and Telecommunications Research Institute
Daejeon, Korea
{bj_park, cleta4u, shk1028, chuh, and machung}@etri.re.kr

Jin-Hun Sohn
Department of Psychology/Brain Research Institute
Chungnam National University
Daejeon, Korea
jhsohn@cnu.ac.kr

***Abstract*— Recently in HCI research, emotion recognition is one of the core processes to implement emotional intelligence. There are many studies using physiological signals in order to recognize human emotions. The purpose of this study is to recognize emotions using autonomic nervous system responses induced by three different emotions (boredom, pain and surprise). Three different emotional states are evoked by emotional stimuli, physiological signals (EDA, ECG, PPG and SKT) for the induced emotions are measured as the reactions of stimuli, and 27 features are extracted from their physiological signals for emotion recognition. The stimuli are used to provoke emotions and tested their appropriateness and effectiveness. Audio-visual film clips used as stimuli are captured originally from movies, documentary, and TV shows with the appropriateness of 86%, 97.3% and 94.1% for boredom, pain and surprise, respectively, and the effectiveness of 5.23 for happiness, 4.96 for pain and 6.12 for surprise (7 point Likert scale). Also, for the three emotion recognition, we propose a Fuzzy c-means clustering based neural networks using the physiological signals. The proposed model consists of three layers, namely, input, hidden and output layers. Here, fuzzy c-means clustering method, two types of polynomial and linear combination function are used as a kernel function in the input layer, the hidden layer and the output layer of neural networks, respectively. To evaluate the performance of emotion recognition of the proposed model, we use the 10-fold cross validation and a comparative analysis shows that the proposed model exhibit higher accuracy when compared with some other models that exist in the literature.**

Keywords-emotion; recognition; stimuli; physiological signal; autonomic nervous system responses; neural networks.

I. INTRODUCTION

Emotion recognition in human-computer interaction (HCI) studies is the one of topic that researcher are most interested in. To recognize human's emotions and feelings, various physiological signals have been widely used to classify emotion [1] because signal acquisition by non-invasive sensors is relatively simple and physiological responses induced by emotion are less sensitive in social and cultural difference [2]. Physiological signal may happen to artifact due to motion, and have difficulty mapping emotion-specific responses pattern, but this has some advantages which are less affected by environment than any other modalities as well as possible to observe user's state in real time. In addition, they also can be acquired spontaneous emotional responses and not caused by responses to social masking or factitious emotion expressions. Finally, measurements of emotional responses by multi-channel physiological signals offer more information for emotion recognition, because physiological responses are related to emotional state [3]. Various physiological signals offer a great potential for the recognition of emotions in computer systems, in order to fully exploit the advantages of physiological measures, standardizations of experimental methods have to be established on the emotional model, stimulus used for the identification of physiological patterns, physiological measures, parameters for analysis, and model for pattern recognition and classification [4].

The objective of this study is to achieve emotion dataset including physiological signals for three emotions (boredom, pain and surprise) induced by emotional stimuli and to develop a recognizer based on neural networks, namely, fuzzy c-means clustering based neural networks.

To induce boredom, pain and surprise emotions, we use audio-visual film clip captured originally from movies, documentary, and TV shows. Electrodermal activity (EDA), skin temperature (SKT), electrocardiac activity (ECG) and photoplethysmography (PPG) are recorded from 200 undergraduate students; during they are exposed to visual-acoustic emotional stimuli. And participants classified their present emotion and assessed its intensity on the emotion assessment scale. As the results of emotional stimulus evaluation, emotional stimuli were shown to mean 92.5% of appropriateness and 5.43 of effectiveness; this means that each emotional stimulus caused its own emotion quite effectively. Also, 27 features are extracted from their physiological signals for emotion recognition.

In order to recognize three emotions, we propose a fuzzy c-means clustering based neural networks (FNN) with polynomial function as a new methodology of emotion recognizer. Neural networks (NNs) have been widely used to deal with pattern recognition problems. It has been shown that the NNs can be trained to approximate complex discriminant functions [5]. NN classifiers can deal with numerous multivariable nonlinear problems for which an

 ISBN: 978-1-61208-250-9

accurate analytical solution is difficult to derive or does not exist [6]. It is found however, that the quality and effectiveness of NN classifiers depend on several essential parameters whose values are crucial to the accurate predictions of the properties being sought. The appropriate NN architecture, the number of hidden layers and the number of neurons in each hidden layer are the important design issues that can immediately affect the accuracy of the prediction. Unfortunately, there is no direct method to identify these factors as they need to be determined on an experimental basis [6]. In addition, it is difficult to understand and interpret the resulting NNs. These difficulties increase once the number of variables and the size of the networks start to increase [7][8]. To alleviate the problems highlighted above, we propose to consider fuzzy c-means clustering based neural networks (FNN) with polynomial function exploiting a direct usage of Fuzzy C-Means clustering and involving polynomials in the description of relationships of the models. The proposed FNN model consists of three layers, namely, input, hidden and output layers. A neuron in input layer employs fuzzy c-means clustering method as a kernel function, namely, this layer relates to the partition function of input space using FCM clustering. Two types of polynomial and linear combination of input signals are used as a kernel function of a neuron in the hidden layer and the output layer of neural networks, respectively. Using two types of polynomial functions can help to improve the characteristic of basic neural networks recognizer and caries out the presentation of a partitioned local space. The proposed FNN recognizer generates a nonlinear discernment function in the output space and has the better performance of emotion recognition. The proposed recognizer is applied to the obtained emotional physiological signals with 27 features and its results are compared with performance of C4.5, SOM (Self-organizing map), Naïve Bayes and SVM (support vector machines) [9]-[13]. The study is organized as follows. In Section II, we discuss emotion stimuli and measurements of physiological signals obtained as autonomic nervous system responses by three emotions. A structure of the proposed FNN is presented with an overall description of a detailed design methodology in Section III. In Section IV, we report on results of emotion recognition for the proposed FNN and conventional models. Finally, concluding remarks are covered in Section V.

II. EMOTIONS AND AUTONOMIC NERVOUS SYSTEM RESPONSES

This section reports a presentation of emotional stimuli and the measurements of physiological signals of autonomic nervous system responses for boredom, pain and surprise emotions

A. Subjects

In this experiment, 200 college students (mean age: 21.7 ± 2.3 years) have participated. They have reported no history of medical illness due to heart disease, respiration, or central nervous system disorder or psychotropic medication. They were introduced to the experiment protocols and filled out a written consent before the beginning of experiment. Also, they were paid $30 USD per session to compensate for their participation.

Figure 1. Example of emotional stimuli

B. Emotional Stimuli

For the three emotions, emotional stimuli are selected the 2-4 min long audio-visual film clips, which are captured originally from movies and TV shows, and provoking pain. Figure 1 is the example of emotional stimuli using audio-visual film clips. The stimulus, which provokes pain, is that it's the more pressure an experimenter put on it after wearing a blood pressure cuff on subjects' arm during 1 minute. The boring stimulus is the combination a presentation of "+" symbol on screen and a repetitive sound of number from 1 to 10 during 3 minutes. The surprise provoking stimulus is the sudden presentation of above images and hog-caller, sound of breaking glass, and thunder during concentration on task during 1 minute. Audio-visual film clips have widely used because these have the desirable properties of being readily standardized, involving no deception, and being dynamic rather than static. They also have a relatively high degree of ecological validity, in so far as emotions are often evoked by dynamic visual and auditory stimuli that are external to the individual [14]-[17].

The audio-visual film clips have been used to provoke emotion and tested their appropriateness and effectiveness. The appropriateness of emotional stimuli means a consistency between the intended emotion by experimenter and the participants' experienced emotion. The effectiveness is an intensity of emotions that participants rated on a 1 to 7 point Likert-type scale, that is, 1 being "least boring" and 7 being "most boring". The appropriateness and effectiveness of these stimuli were as follows; appropriateness and effectiveness of boredom were 86.0% and 5.23±1.36, in pain 97.3% appropriateness and 4.96±1.34 effectiveness and 94.1% appropriateness and 6.12±1.14 effectiveness in surprise.

C. Measurements of Physisological Signals and Feature extraction

To collect physiological signals for three emotions, the laboratory is a room with 5m x 2.5m size having a sound-proof (lower than 35dB) of the noise level where any outside noise or artifact are completely blocked. A comfortable chair is placed in the middle of the laboratory and 38 inch TV

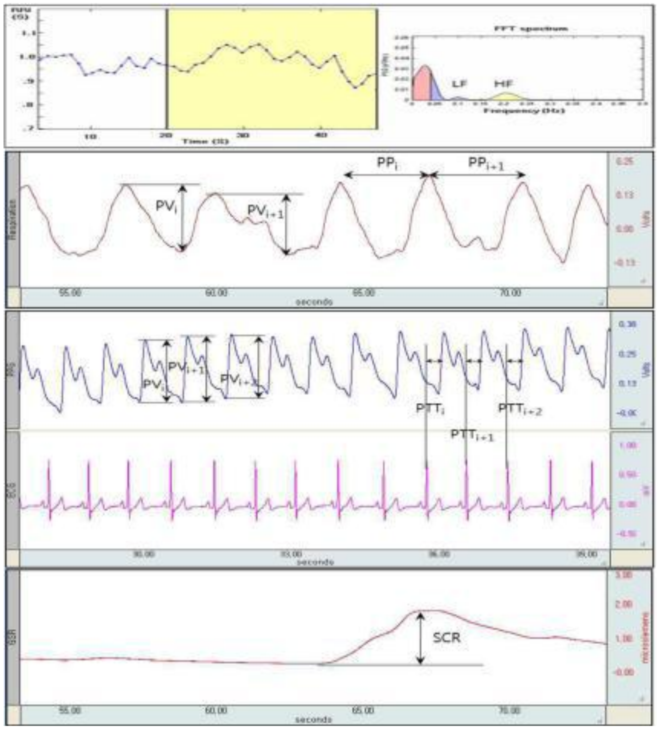


Figure 2. Physiological signals and feature extraction

monitor set for presentation of film clips is placed in front of the chair. An intercommunication device is placed to the right side of chair for subjects to communicate with an experimenter. A CCTV is installed on the top of the monitor set to observe participant's behaviors and their behaviors are storage through the monitor and a video cassette recorder outside the laboratory.

Prior to the experiment, subjects are introduced to detail experiment procedures and have an adaptation time to feel comfortable in the laboratory setting. Then they are attached electrodes on their wrist, finger, and ankle for measurement of physiological signals. Physiological signals are measured for 60 sec prior to the film clip presentation (baseline) and for 2 to 4 min during the presentation of the film clips (emotional state), then for 60 sec after presentation of the film clips as recovery term. Subjects rate the emotion that they experienced during presentation of the film clip on the emotion assessment scale.

The dataset of physiological signals such as skin temperature (SKT), electrodermal activity (EDA), photoplethysmography (PPG), and electrocardiogram (ECG) are collected by MP150 Biopac system Inc. (USA). SKT electrodes are attached on the first joint of non-dominant ring finger and on the first joint of the non-dominant thumb for PPG. EDA is measured with the use of 8 mm AgCl electrodes placed on the volar surface of the distal phalanges of the index and middle fingers of the non-dominant hand. Electrodes are filled with a 0.05 molar isotonic NaCl paste to provide a continuous connection between the electrodes and the skin. ECG electrodes are placed on both wrists and one left ankle with two kinds of electrodes, sputtered and AgCl ones. The left-ankle electrode is used as a reference.

The signals are acquired for 1 minute long baseline state prior to presentation of emotional stimuli and 1-2 minutes long emotional states during presentation of the stimuli as emotional state. To extract features, the obtained signals are analyzed for 30 seconds from the baseline and the emotional state by AcqKnowledge (Ver. 3.8.1) software (USA) as shown in Fig. 2. 27 features, namely, bSCL, bSCR, eSCL, eSCR, dSCL, and dSCR features from EDA signal, BmeanSKT, EmeanSKT and DmeanSKT from SKT, bHR, bLF, bHF, bHRV, eHR, eLF, eHF, eHRV, dHR, dLF, dHF and dHRV from ECG, and bBVP, bPPT, eBVP, ePPT, dBVP and dPPT from PPG are extracted from the obtained emotional physiological signals. These are the features that are commonly used in physiological signal analysis.

III. Structure and Development of Fuzzy c-means Clustering based Neural Networks

Neural Network (NN) is a computational intelligence model inspired by the structure and functional aspects of biological neurons [11]. NN has been widely used to deal with pattern recognition problems. The generic topology of NN consists of three layers as shown in Fig. 1. A neuron in the input layer is connected to a layer of hidden neuron, which is connected to output neuron. The activity of the input neurons represents the raw information that is fed into the network, the activity of each hidden neuron is determined by the activities of the input neuron and the weights on the connections between the input and the hidden, and the behavior of the output depends on the activity of the hidden neurons and connection weights between the hidden and the output layer.

The proposed Fuzzy c-means clustering based neural networks (FNN) exhibit a similar topology as the one encountered in simple neural network. However the functionality and the associated design process exhibit some evident differences. In particular, the kernel fields of neurons do not assume any explicit functional form (say, Gaussian, ellipsoidal, etc.), but are directly reflective of the nature of the data and come as the results of fuzzy clustering in input layer, polynomials in hidden layer and linear combination of

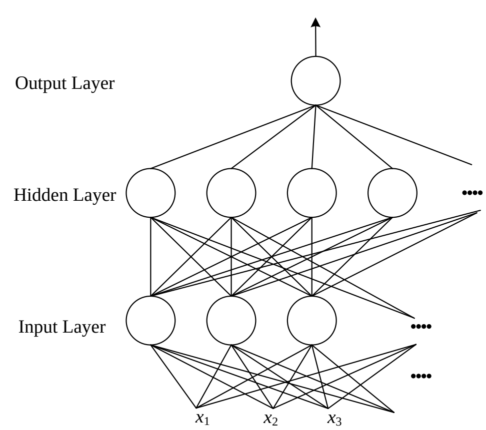


Figure 3. General structure of neural networks

activation signals in output layer. These networks results as a synergy between two other general constructs such as fuzzy c-means clustering [18][19] and neural networks [11].

In input layer, considering the prototypes v_1, v_2, …, v_c formed by the fuzzy c-means (FCM) clustering method, the kernel of a neuron is expressed in the following way

$$IK_i(\mathbf{x}) = 1 / \sum_{j=1}^{c} \left(\frac{\|\mathbf{x}-\mathbf{v}_i\|^2}{\|\mathbf{x}-\mathbf{v}_j\|^2} \right) \tag{1}$$

The FCM clustering method comes as a standard mechanism aimed at the formation of 'c' fuzzy sets. The objective function Q guiding this clustering is expressed as a sum of the distances of individual data from the prototypes v_1, v_2, …, and v_c,

$$Q = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{m} \|\mathbf{x}_k - \mathbf{v}_i\|^2 \tag{2}$$

Here, $\| \|$ denotes a certain distance function; 'm' stands for a fuzzification factor (coefficient), m=2.0. N is the number of patterns. The resulting partition matrix is denoted by $\mathbf{U}=[u_{ik}]$, i=1, 2,…,c; k=1, 2, …, N. While there is a substantial diversity as far as distance functions are concerned, here we adhere to a weighted Euclidean distance taking on the following form

$$\|\mathbf{x}_k - \mathbf{v}_i\|^2 = \sum_{j=1}^{n} \frac{(x_{kj} - v_{ij})^2}{\sigma_j^2} \tag{3}$$

with σ_j being a standard deviation of the j-th variable. While not being computationally demanding, this type of distance is still quite flexible and because of that commonly used.

The minimization of Q is realized in successive iterations by adjusting both the prototypes and entries of the partition matrix, that is min $Q(\mathrm{U}, v_1, v_2, \ldots, v_c)$. The resulting partition matrix becomes output signals of kernel function in input layer, namely, U=IK.

In addition, the kernel functions of hidden layer come in the product of polynomials of inputs and activation signals which are output signals of input layer.

$$HK_i = f_i(\mathbf{x}) \cdot IK_i(\mathbf{x}) \tag{4}$$

Here, $f_i(\mathbf{x})$ is a certain polynomial coming from the following family of relationships

$$\begin{aligned} &\text{Constant}: \; f_i(\mathbf{x}) = a_{i0} \\ &\text{Linear}: \; f_i(\mathbf{x}) = a_{i0} + \sum_{j=1}^{n} a_{ij} x_j \end{aligned} \tag{5}$$

These functions are activated by partition matrix and lead to local regression models. The proposed FNNs governed by (5) are furnished with a great deal of degrees of freedom which may result with improved their emotion recognition rates when contrasted with conventional models.

The neuron located at the output layer completes a linear combination of the activation levels of the hidden layer.

$$\begin{aligned} y &= \sum_{i=1}^{c} HK_i = \sum_{i=1}^{c} f_i(\mathbf{x}) \cdot IK_i(\mathbf{x}) \\ &= \sum_{i=1}^{c} f_i(\mathbf{x}) \cdot \left(1 / \sum_{j=1}^{c} \left(\frac{\|\mathbf{x}-\mathbf{v}_i\|^2}{\|\mathbf{x}-\mathbf{v}_j\|^2} \right) \right) \end{aligned} \tag{6}$$

y is a representation of FNNs as a discriminant function.

IV. RESULTS OF EMOTION RECOGNITION

Our objective is to quantify the performance of the proposed FNNs recognizer and to compare it with the performance of some other recognizer reported in the literature. In the assessment of the performance of the recognizer, we use the recognition accuracy for three emotions. The experiments completed in this study are reported for the 10 fold cross-validation for assessing how the results of a statistical analysis will generalize to an independent dataset.

For the recognition of boredom, pain and surprise emotions, Table I contrasts the recognition accuracy (%) of the proposed FNNs with other well-known methods (C4.5, SOM, Naïve Bayes and SVM) studied in the literatures [9]-[14]. The experimental results reveal that the proposed approach and the resulting model outperform the existing methods in terms of better prediction capabilities on feature space.

C4.5 developed by R. Quinlan is one of statistical classifier algorithms and builds decision trees from a set of training data. C4.5 starts with large sets of cases belonging to known classes. The cases, described by any mixture of nominal and numeric properties, are scrutinized for patterns that allow the classes to be reliably discriminated. These patterns are then expressed as models, in the form of decision trees or sets of if-then rules, which can be used to classify new cases, with emphasis on making the models understandable as well as accurate.

SOM (self-organizing map), called Kohonen map, is a type of artificial neural networks in the unsupervised learning category and generally present a simplified, relational view of a highly complex data set. This is called a topology-preserving map because there is a topological structure imposed on the nodes in the network. A topological map is simply a mapping that preserves neighbourhood relations. The goal of training is that the "winning" unit in the target space is adjusted so that it is more like the particular pattern. Others in the neighbourhood of output are also adjusted so that their weights more nearly match that of the input pattern.

The Naïve Bayes algorithm is a classification algorithm based on Bayes rule and particularly suited when the dimensionality of the inputs is high. When the dependency relationships among the features used by a classifier are

 ISBN: 978-1-61208-250-9

unknown, we generally proceed by taking the simplest assumption, namely, that the feature are conditionally independent given the category, that is,

$$p(\omega_k \mid \mathbf{x}) \propto \prod_{i=1}^{d} p(x_i \mid \omega_k) \quad (7)$$

This so-called naïve Bayes rule often works quite well in practice, and it can be expressed by a very simple belief net.

SVM (support vector machine) finds a hyperplane based on support vector to analyze data and recognize patterns. The complexity of the resulting recognizer is characterized by the number of support vectors rather than the dimensionality of the transformed space. The goal in training SVM is to find the separating hyperplane with the largest margin. We expect that the larger the margin, the better generalization of the recognizer.

The more detail results of emotion recognition accuracy by each algorithm are like from Table II to VIII. In analysis of C4.5, accuracy of each emotion has range of 42% to 74%. Boredom is recognized by C4.5 with 73.9%, pain 42.0%, and surprise 58.9% as shown in Table II. In Table III and IV, the results of Naïve Bayes and SVM are shown for three emotions, respectively. Naïve Bayes provides accuracy of 71.88 % when it recognized all emotions. In boredom, accuracy of 73.9% is achieved, 71.6% in pain, 66.7% in surprise. The results of emotion recognition using SVM show recognition accuracy of 67%, 62.1%, and 57.3% according to orders of boredom, pain, and surprise. For the supervised and unsupervised SOM, 61.45% and 59.22% are obtained, respectively, as recognition accuracy for three emotions. For the each emotion, boredom, pain and surprise, 69.3%, 52.7% and 62% are obtained in supervised SOM and 68.2%, 56.8%, and 53.1% are resulted from unsupervised SOM. Finally, the recognition accuracy of the proposed FNN with constant and linear has 70.2% and 74.49%, respectively for three emotion recognition. In case of constant, each emotion is recognized by FNN with 79% of boredom, 61.5% of pain and 69.8% of surprise. 72.2%, 74.6% and 76.6% are shown as results of emotion recognition by FNN with linear for boredom, pain and surprise, respectively.

C4.5 and Naïve Bayes were coming from the Classification Toolbox of MATLAB. For SVM, we have used Duda's Toolbox (www.yom-tov.info/toolbox.html). SOM toolbox available in MATLAB has offered SOM algorithms, see www.cis.hut.fi/projects/somtoolbox/.

TABLE I. RESULT OF EMOTION RECOGNITION

Models		Accuracy (%)
C4.5		58.47
Naïve Bayes		71.88
SVM		62.01
SOM	Supervised	61.45
	Unsupervised	59.22
Proposed FNNs	Constant	70.2
	Linear	74.49

TABLE II. RESULT OF EMOTION RECOGNITION BY C4.5

	Boredom	Pain	Surprise
Boredom	73.9	2.3	23.9
Pain	22.5	42	35.5
Surprise	28.6	12.5	58.9

TABLE III. RESULT OF EMOTION RECOGNITION BY NAÏVE BAYES

	Boredom	Pain	Surprise
Boredom	77.8	2.3	19.3
Pain	8.3	71.6	19.5
Surprise	16.1	16.7	66.7

TABLE IV. RESULT OF EMOTION RECOGNITION BY SVM

	Boredom	Pain	Surprise
Boredom	67	8	25
Pain	5.9	62.1	32
Surprise	14.1	28.6	57.3

TABLE V. RESULT OF EMOTION RECOGNITION BY SUPERVISED SOM

	Boredom	Pain	Surprise
Boredom	69.3	10.2	20.5
Pain	3	52.7	44.4
Surprise	7.8	30.2	62

TABLE VI. RESULT OF EMOTION RECOGNITION BY UNSUPERVISED SOM

	Boredom	Pain	Surprise
Boredom	68.2	7.4	19.9
Pain	8.9	56.8	32.5
Surprise	14.1	30.7	53.1

TABLE VII. RESULT OF EMOTION RECOGNITION BY FNN WITH CONSTANT

	Boredom	Pain	Surprise
Boredom	79	4.5	16.5
Pain	4.1	61.5	34.3
Surprise	9.4	20.8	69.8

TABLE VIII. RESULT OF EMOTION RECOGNITION BY FNN WITH LINEAR

	Boredom	Pain	Surprise
Boredom	72.2	4	23.9
Pain	3	74.6	22.5
Surprise	4.2	19.3	76.6

V. CONCLUSION

In this study, we have discussed the acquisition of emotional physiology signals using emotion stimuli and the design of an emotion recognition methodology for boredom, pain and surprise emotions. The emotion stimuli used to induce a participant's emotion were evaluated for their suitability and effectiveness. The result showed that emotional stimuli have the appropriateness of 92.5% and the effectiveness of 5.4 point (7 point Likert scale) on average. Twenty seven features have been extracted by means of the statistical and the geometric approaches in time and frequency domain from physiological signals such as EDA, SKT, PPG and ECG. These signals have been induced by emotional stimuli. In order to recognize the three emotions with physiological signals, we have proposed fuzzy c-means clustering based neural networks (FNNs) with three layers. Fuzzy c-means clustering method as a kernel function has used in a neuron in input layer. A kernel function of a neuron in the hidden layer used two types of polynomial functions and the linear combination of input signals were used in output layer of the proposed networks. Using two types of polynomial functions helped to improve the characteristic of a recognizer and carried out the presentation of a partitioned local space. The proposed FNN recognizer generates a nonlinear discernment function in the output space and has the better performance of emotion recognition. As shown in results, the proposed FNN with linear function has recognition accuracy 74.49% for the three emotions. The proposed recognizer will lead to better chance to recognize human emotions by using physiological signals in the emotional interaction between man and machine.

ACKNOWLEDGMENT

This research was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (No. 2012K001330 & 2012K001339).

REFERENCES

[1] J.Wagner, J. Kim, and E. Andre, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," IEEE International Conference on Multimedia and Expo, pp. 940-943, 2005.

[2] N. Sebe, I. Cohen, and T.S. Huang, Multimodal emotion recognition, in Handbook of Pattern Recognition and Computer Vision, Amsterdam: Publications of the Universiteit van Amsterdam, 2005, PP. 1-23.

[3] P.D. Drummond, and S.-H. Quah, "The effect of expressing anger on cardiovascular reactivity and facial blood flow in Chinese and Caucasians," Psychophysiology, pp. 190-196, 2001.

[4] R.O. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," IEEE Transaction on Pattern Analysis and Machine Intelligence, pp. 1175-1191, 2001.

[5] R. P. Lippman, "An introduction to computing with neural nets," IEEE ASSP Magazine, vol. 4, no. 2, pp. 4-22, 1981.

[6] Y. Al-Assaf, and H. El Kadi, "Fatigue life prediction of composite materials using polynomial classifiers and recurrent neural networks," Composite Structures, vol. 77, pp. 561-569, 2007.

[7] F. Ros, M. Pintore, and J.R. Chretien, "Automatic design of growing radial basis function neural networks based on neighborhood concepts," Chemometrics and Intelligent Laboratory Systems, vol. 87, pp. 231-240, 2007.

[8] C. Zhang, J. Jiang and M. Kamel, "Intrusion detection using hierarchical neural networks," Pattern Recognition Letters, vol. 26, pp. 779-791, 2005.

[9] J.R. Quinlan, C4.5 Programs for Machine Learning, San Mateo, CA: Morgan Kaufmann, 1992.

[10] C. Lim, and J.-H. Chang, "Adaptive Kernel Function of SVM for Improving Speech/Music Classification of 3GPP2 SMV", ETRI Journal, vol. 33, no. 6, pp.871-879, 2011.

[11] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern classification, 2nd ed. Wiley-Interscience, New York, 2000.

[12] D. Wasserman, Advanced Methods in Neural Computing, New York, Van Nostrand Reinhold, 1993 pp. 35–55.

[13] T. Kohonen, "Self-Organizing Maps," Proceedings of the IEEE, vol. 79, pp. 1464-1480, 1990.

[14] J.J. Gross, and R.W. Levenson, "Emotion elicitation using films," Cognition and Emotion, vol. 9, pp. 87-108, 1995.

[15] R.S. Lazarus, J.C. Speisman, A.M. Mordkoff, and L.A. Davidson, "A Laboratory study of psychological stress produced by an emotion picture film," Psychological Monographs, vol. 76, pp. 553, 1962.

[16] M.H. Davis, J.G. Hull, R.D. Young, and G.G. Warren, "Emotional reactions to dramatic film stimuli: the influence of cognitive and emotional empathy," Journal of personality and social psychology, vol. 52, pp. 126-133, 1987.

[17] D. Palomba, M. Sarlo, A. Angrilli, A. Mini, and L. Stegagno, "Cardiac responses associated with affective processing of unpleasant film stimuli," International Journal of Psychophysiology, vol. 36, pp. 45-57, 2000.

[18] A. Aiyer, K. Pyun, Y.Z. Huang, D.B. O'Brien and R.M. Gray, "Lloyd clustering of Gauss mixture models for image compression and classification," Signal Processing: Image Communication, vol. 20 pp. 459-485, 2005.

[19] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

 ISBN: 978-1-61208-250-9

Classification of Human Emotions from Physiological signals using Machine Learning Algorithms

Recognition of Pain, Boredom, and Surprise Emotions

Eun-Hye Jang, Byoung-Jun Park, Sang-Hyeob Kim, Myoung-Ae Chung
BioHeath IT Convergence Research Department, IT Convergence Services Future Technology Research Department
Electronics and Telecommunications Research Institute
Daejeon, S. Korea
cleta4u@etri.re.kr; bj_park@etri.re.kr; shk1028@etri.re.kr; machung@etri.re.kr

Mi-Sook Park, Jin-Hun Sohn
Department of Psychology/Brain Research Institute
Chungnam National University
Daejeon, S. Korea
peaceful1011@nate.com ; jhsohn@cnu.ac.kr

***Abstract*—Emotion recognition is one of the key steps towards emotional intelligence in advanced human-machine interaction. Recently, emotion recognition using physiological signals has been performed by various machine learning algorithms as physiological signals are important for emotion recognition abilities of human-computer systems. The purpose of this study is to classify three different emotional states (boredom, pain, and surprise) from physiological signals using several machine learning algorithms and to identify the optimal algorithms being able to classify these emotions. 217 subjects participated in this experiment. The emotional stimuli designed to induce three emotions (boredom, pain, and surprise) were presented to subjects and physiological signals were measured for 1 minute as baseline and for 1-1.5 minutes during emotional states. The obtained signals were analyzed for 30 seconds from the baseline and the emotional state and 27 parameters were extracted from these signals. For classification of three different emotions, machine learning algorithms of Decision tree, k-NN (k-nearest neighbor algorithm), LDA (linear discriminant analysis), and SVM (support vector machine) were done by using the difference values of signal parameters subtracting baseline from the emotional state. Classification accuracy using LDA was 74.9% and the result of emotion recognition using Decision Tree showed that accuracy to recognize all emotions was 67.8%. In analysis of k-NN and SVM, classification accuracy was 62.0%. The result of emotion recognition shows that LDA is the best algorithm being able to classify pain, surprise, and boredom emotions. This led to better chance to recognize other emotions except human basic emotions and to assist more accurate and greater understanding on emotional interactions between man and machine based on physiological signals.**

Keywords-emotion; pain; surprise; boredom; physiological signals; machine learning algorithm

I. INTRODUCTION

Recently, in an attempt to categorize and understand human emotions, psychologists and engineers have tried to analyze various modalities such as facial expressions, voices, gestures, and physiological signals [1]. In particular, various physiological signals have been widely used to recognize human emotions for the following advantages. Although physiological signal may happen to artifact due to motion or other environmental factors, its signal acquisition by non-invasive sensors is relatively simple and it is possible to observe user's state in real time. Also, physiological responses can be acquired spontaneous emotional responses not by responses to social masking or factitious emotion expressions and are less sensitive in social and cultural difference [2]. Finally, various physiological signals offer more information for emotion recognition, because physiological responses are related to emotional state [3] and are considered a great potential for emotion recognition in computer systems.

Many emotion researches have studied physiological signals induced by basic emotions [4-12] and recently, emotion recognition based on physiological signals was performed by various algorithms. Studies on emotion classification from physiological responses using machine learning algorithms (e.g., Fisher projection, k-nearest neighbor algorithm, and support vector machines, etc.) have mainly focused on responses induced by basic emotions such as happiness, sadness, anger, fear, and disgust [13-17]. On the other hand, other emotions such as boredom, pain etc. have been least investigated and there are little results of emotion classification on these emotions. Although these emotions aren't basic emotion, they are emotion that human have often experienced in real life and it is needed to classify them from multi-channel physiological signals using machine learning algorithms.

The purposes of this study are to classify three different emotions (pain, boredom, and surprise) using multi-channel physiological signals (ECG, EDA, PPG, and SKT) and to identify the optimal algorithms being able to recognize them. We have operationally defined that surprise emotion is 'startle' response to a sudden unexpected stimulus such as a

 ISBN: 978-1-61208-250-9

flash of light, a loud noise, or a quick movement near the face [17-18]. For emotion classification, there are used Decision Tree (which is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences), k-NN (k-nearest neighbor algorithm, which is a method for classifying objects based on closest training examples in the feature space), LDA (linear discriminant analysis, which is a method used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events), and SVM (support vector machine, supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis).

II. EXPERIMENTAL METHODS

A. Subjects

129 college students (60 males, 69 females, ages 22.0± 2.2 years) and 88 high school students (37 males, 51 females, ages 16±1.3 years) participated in this experiment. They were normal persons who reported no history of medical illness due to heart disease, respiration, or central nervous system disorder. They were introduced to the experiment protocols and filled out a written consent before the beginning of experiment. Also, they were paid $30 USD per session to compensate for their participation.

B. Emotional Stimuli

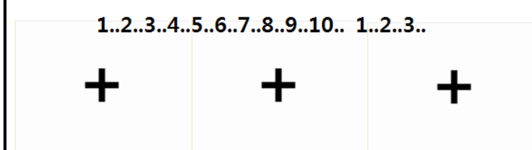
The emotional stimuli used in experiment, which are the 1-3 min long audio-visual stimuli and stimulus provoking pain, had been demonstrated their appropriateness and effectiveness by preliminary psychometric experiment. The appropriateness of emotional stimuli means a consistency between the intended emotion by experimenter and the participants' experienced emotion. The effectiveness is an intensity of emotions that participants rated on a 1 to 7 point Likert-type scale (e.g.., 1 being "least boring" and 7 being "most boring"). The apporiateness and effectiveness of these stimuli were as follows; appropriateness and effectiveness of pain were 97.3% appropriateness and 4.96 ± 1.34 effectiveness, in boredom were 86.0% and 5.23±1.36, and 94.1% appropriateness and 6.12±1.14 effectivess in surprise.

The example of each emotion stimulus is like Table I. The pain provoking stimulus is that it's the more pressure an experimenter put on it after wearing a blood pressure cuff on subjects' arm during 1 minute. The boring stimulus is the combination a presentation of "+" symbol on screen and a repetitive sound of number from 1 to 10 during 3 minutes. The surprise provoking stimulus is the sudden presentation of above images and hog-caller, sound of breaking glass, and thunder during concentration on task during 1 minute.

C. Experimental Settings and Procedures

The laboratory is a room with 5m x 2.5m size having a sound-proof (lower than 35dB) of the noise level where any outside artifact are completely blocked. A comfortable chair is placed in the middle of the laboratory and TV monitor set for presentation of film clips is placed in front of the chair. An intercommunication device is placed to the right side of chair for subjects to communicate with an experimenter. A CCTV is installed on the top of the monitor set to observe participant's behaviors and their behaviors are storage through the monitor and a video cassette recorder outside the laboratory.

TABLE I. THE EXAMPLE OF EMOTION STIMULI

Emotion	Stimulus
pain	
	Induction of pain using blood pressure cuff (1 min)
boredom	
	Repetitive sounds of number from 1 to 10 (3 min)
surprise	
	Sudden presentation of above images and hog-caller, sound of breaking glass, and thunder during concentration on task (1 min)

Prior to the experiment, subjects are introduced to detail experiment procedures and have an adaptation time to feel comfortable in the laboratory setting. Then they are attached electrodes on their wrist, finger, and ankle for measurement of physiological signals. Physiological signals are measured for 1 minute prior to the emotional stimuli (baseline) and for 1 to 3 minnutes during the presentation of stimuli (emotional state), then for 1 minute after presentation of the emotional stimuli as recovery term. Subjects rated the own emotion that they experienced during emotional state (Fig. 1).

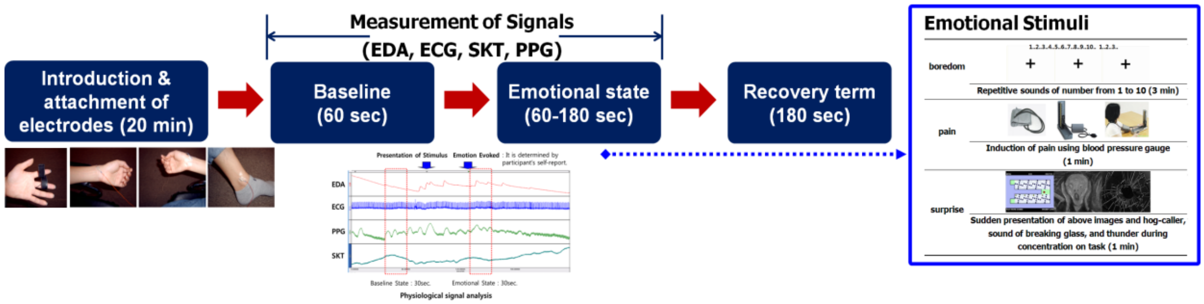
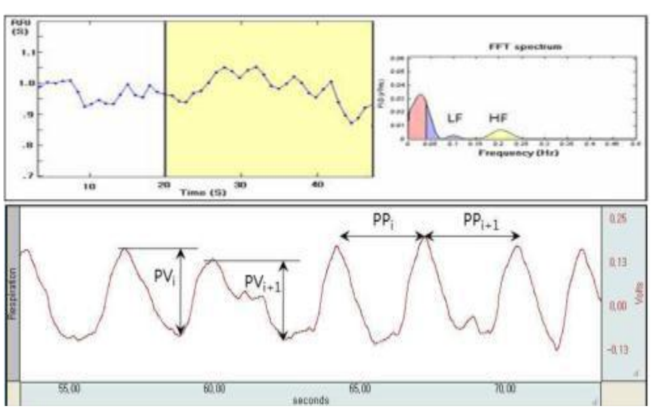


Figure 1. Experimental procedure.

D. *Measuremets of Physisological Signals and Feature extraction*

The dataset of physiological signals, electrocardiogram (ECG), electrodermal activity (EDA), skin temperature (SKT), and photoplethysmography (PPG) were collected by MP150 Biopac system Inc. (USA). For measurement of ECG, ECG electrodes were placed on both wrists and one left ankle with two kinds of electrodes, sputtered and AgCl ones. The electrode on left-ankle was used as a reference. EDA was measured with the use of 8 mm AgCl electrodes placed on the volar surface of the distal phalanges of the index and middle fingers of the non-dominant hand. The electrodes were filled with a 0.05 molar isotonic NaCl paste to provide a continuous connection between the electrodes and the skin. SKT electrode was attached on the first joint of the non-dominant ring finger and on the first joint of the non-dominant thumb for PPG. These signals were sampled with sampling rate 250Hz, and appropriate amplification and band-pass filtering were performed.

The signals are acquired for 1 minute long baseline state prior to presentation of emotional stimuli and 1-3 minutes long emotional states during presentation of the stimuli as emotional state. To extract features, the obtained signals are analyzed for 30 seconds from the baseline and the emotional state by AcqKnowledge (Ver. 3.8.1) software (USA) as shown in Fig. 2. 27 features are extracted and analyzed from the obtained physiological signals (Table II).



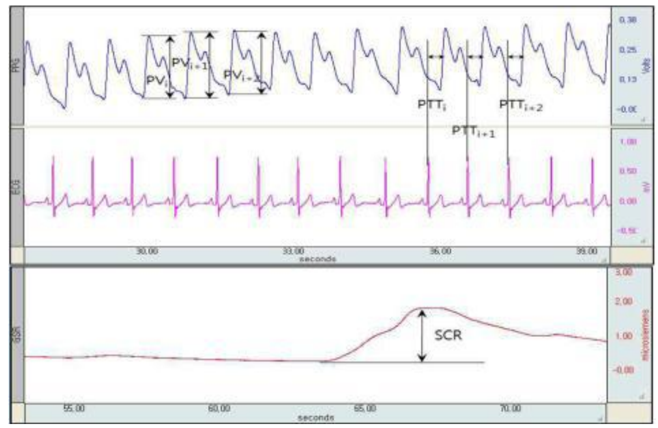


Figure 2. The example of acquired physiological signals.

TABLE II. THE EXTRACTED PHYSIOLOGICAL FEATURES

Signals	Features
EDA	b_SCL, b_SCR, e_SCL, e_SCR, d_SCL, d_SCR
SKT	b_meanSKT, e_meanSKT, d_meanSKT
PPG	b_BVP, b_PPT, e_BVP, e_PPT, d_BVP, d_PPT
ECG	b_HR, b_LF, b_HF, b_HRV, e_HR, e_LF, e_HF, e_HRV, d_HR, d_LF, d_HF, d_HRV

b_: baseline
e_: emotional state
d_: ‘e_’ – ‘b_’

E. *Machine Learning Algorithms for Emotion Recognition*

For three different emotion classification, 4 machine learning algorithms, Dicision tree, k-NN, LDA, and SVM were applicated by using the extracted features. Decision tree is a hierarchy based classifier in which each branch node represents an option between a number of alternatives, and each leaf node represents a decision and a decision support tool that uses a tree-like graph or model of decisions and their possible consequences[19]. It can select from among a large number of variables those and their interactions that are most important in determining the outcome variable to be explained. Given the data represented at a node, either declare that node to be a leaf (and state what category to assign to it), or find another property to use to split the data into subsets. Decision trees have various advantages: it is possible to validate a model using statistical tests and performs well with large data in a short time. But, decision tree learners can create over-complex trees that do not generalise the data well.

K-NN is a method for classifying objects based on closest training examples in the feature space. It is a method for classifying objects based on closest training examples in the feature space and is a simple and efficient classifier, so it is proper to apply KNN to emotion recognition. The k-nearest neighbor classifier assigns an utterance to an emotional state according to the emotional state of the k utterances that are closest to uξ in the measurement space. It's a method for classifying patterns based on closest training datasets without probability arguments in the feature space. K-NN decision rule provides a simple nonparametric procedure for the assignment of a class label to the input pattern based on the class labels represented by the k-closest neighbors of the vector. However, the disadvantages of k-NN is that systematic methods for selecting the optimum number of the closest neigh bors and the most suitable distance measure are hard to find.

LDA which is one of the linear models is a method used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events. LDA finds the direction to project data on so that between-class variance in maximized and within-class variance in minimized, and then offers a linear transformation of predictor variables which provides a more accurate discrimination [20]. In LDA, the measurement space is transformed so that the separability between the emotional states is maximized. The separability between the emotional states can be expressed by several criteria.

SVM is non-linear model, which are used the well-known emotion algorithms and support vector classifier separates the emotional states with a maximal margin. The advantage of support vector classifier is that it can be extended to nonlinear boundaries by the kernel trick. SVM supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. SVM is designed for two class classification by finding the optimal hyperplane where the expected classification error of test samples is minimized and has utilized as a pattern classifier to overcome the difficulty in pattern classification due to the large amount of within-class variation of features and the overlap between classes, although the features are carefully extracted [20]. The goal in training SVM is to find the separating hyperplane with the largest margin. We expect that the larger the margin, the better generalization of the recognizer [21].

In the next section, we will discuss the compartive results of emotion classification by the four algorithms as the mentioned above. These algorithms are well-known general methods studied in lots of literatures. We have used the Classification Toolbox of MATLAB for Decision tree and Duda's Toolbox, see www.yom-tov.info/toolbox.html, for k-NN, LDA and SVM. We used feature normalization and the related parameters of algorithms used default values, which have offered with toolbox.

III. RESULTS

The purpose of this study is to compare the performance of each classifier and we used the recognition accuracy as the performance of a classifier for three emotions, i.e., pain, boredom, and surprise. The performance of each classifier was evaluated by 10 fold cross-validation for the overfitting problem and the results of this study are reported for those. For the recognition of three emotions, Table III contrasts the recognition accuracy (%) of the used algorithms. Our result showed that the optimal algorithm being able to recognize three emotions was LDA (74.9%).

The more detail results of emotion recognition accuracy by each algorithm are like from Table IV to VII. Decision tree provided accuracy of 67.8% when it recognized all emotions and accuracy of each emotion had range of 58.9% to 76.1%. Pain was recognized by Decision tree with 69.8%, boredom 76.1%, and surprise 58.9% as shown in Table IV. In analysis of k-NN, the accuracy of all emotions was 62.0% and accuracy of each emotion showed that accuracy of 61.5% was achieved in pain, 68.2% in boredom, and 56.8% in surprise. LDA had recognition accuracy of 74.9% in all emotions as shown in Table III. LDA showed recognition accuracy of 76.3%, 75.6%, and 72.9% according to orders of pain, boredom, and surprise. Finally, as can be seen in Table VII, the result of the SVM was 62.0% in all emotions and this algorithm successfully recognized pain (62.1%), boredom (67.0%), and surprise (57.3%).

TABLE III. RESULT OF EMOTION CLASSIFICATION BY MACHINE LEARNING ALGORITHMS

Algorithm	Accuracy (%)	Features (N)
Decision tree	67.8	27
k-NN	62.0	27
LDA	74.9	27
SVM	62.0	27

 ISBN: 978-1-61208-250-9

TABLE IV. RESULT OF EMOTION CLASSIFICATION BY DECISION TREE

	Pain	Boredom	Surprise
Pain	69.2	6.5	24.3
Boredom	6.8	76.1	17.0
Surprise	21.9	19.3	58.9

TABLE V. RESULT OF EMOTION CLASSIFICATION BY K-NN

	Pain	Boredom	Surprise
Pain	61.5	7.7	30.8
Boredom	8.0	68.2	23.9
Surprise	27.6	15.6	56.8

TABLE VI. RESULT OF EMOTION CLASSIFICATION BY LDA

	Pain	Boredom	Surprise
Pain	76.3	1.8	21.9
Boredom	5.7	75.6	18.8
Surprise	20.8	6.3	72.9

TABLE VII. RESULT OF EMOTION CLASSIFICATION BY SVM

	Pain	Boredom	Surprise
Pain	62.1	5.9	32.0
Boredom	8.0	67.0	25.0
Surprise	28.6	14.1	57.3

IV. CONCLUSION

We identified that three different emotions (pain, boredom, and surprise) were classified by machine learning algorithms from various physiological features. For this, twenty seven features were extracted by means of the statistical and the geometric approaches in time and frequency domain from physiological signals i.e., ECG, EDA, SKT, and PPG and these signals were induced by emotional stimuli.

Also, we recognized three emotions by 4 machine learning algorithms of Decision tree, k-NN, LDF, and SVM. Our result showed that LDA is the best algorithm being able to classify these emotions. The LDA algorithm offers many advantages in other pattern recognition tasks such as face recognition or speech recognition etc. LDA finds the vectors in the underlying space that best discriminate among classes. LDA method tries to maximize the between-class differences and minimize the within-class ones. LDA method is good at discriminating different classes because it is a surveillance method. But LDA always suffers from a small sample size problem. The problem will happen when the number of training samples is less than the total number of physiological features. Although LDA method has some problems, we think that our result is reliable and stable because it is based on sufficient sample size of 227 subjects' data and 27 features.

The result of this study could help emotion recognition studies lead to better chance to recognize various human emotions by using physiological signals. Also, this result can be useful in developing an emotion theory, or profiling emotion-specific physiological responses, as well as establishing the basis for emotion recognition system in human-computer interaction. Physiological signals offer a great potential for the recognition of emotions in computer systems. But, in order to fully exploit the advantages of physiological measures, standardization needs to be established on the emotional model, stimulus used for the identification of physiological patterns, physiological measures, parameters for analysis, and model for pattern recognition and classification [22].

Future studies are needed to obtain additional signals from other modalities such as facial expression, face temperature, or voice to improve classification rate. And more research is needed to obtain stability and reliability of this result compare with accuracy of emotion classification using other algorithms.

ACKNOWLEDGMENT

This research was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (No.012K001330 & 2012K001339).

REFERENCES

[1] J.Wagner, J. Kim, and E. Andre, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," *IEEE International Conference on Multimedia and Expo*, pp. 940-943, 2005.

[2] N. Sebe, I. Cohen, and T.S. Huang, Multimodal emotion recognition, in *Handbook of Pattern Recognition and Computer Vision*, Amsterdam: Publications of the Universiteit van Amsterdam, 2005, pp. 1-23.

[3] P.D. Drummond, and S.-H. Quah, "The effect of expressing anger on cardiovascular reactivity and facial blood flow in Chinese and Caucasians," *Psychophysiology*, pp. 190-196, 2001.

[4] O. Alaoui-Ismaili, O. Robin, H. Rada, A. Dittmar, A. and E. Vernet-Maury, "Basic emotions evoked by odorants: comparison between autonomic responses and self-evaluation," *Physiology and Behavior*, 62, pp. 713-720, 1997.

[5] A.R. Ax, "The physiological differentiation between fear and anger in humans," *Psychosomatic Medicine*, vol. 15, pp. 147-150, 1953.

[6] F.A. Boiten, "Autonomic response patterns during voluntary facial action," *Psychophysiology*, vol. 33, pp. 123-131, 1996.

[7] T.C. Kanade, and Y. Tian, "Comprehensive database for facial expression analysis," *Proceeding of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46-53, 2000.

[8] D. Palomba, M. Sarlo, A. Angrilli, and A. Mini, "Cardiac responses associated with affective processing of unpleasant film stimulus," *International Journal of Psychophysiology*, vol. 36, pp. 45-57, 2000.

[9] R.W. Picard, E. Vyzas, and J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1175-1191, 2001.

[10] R. Sinha, and O.A. Parsons, "Multivariate response patterning of fear and anger," *Cognition and Emotion*, vol. 10, pp. 173-198, 1996.

[11] G. Stemmler, *Physiological processes during emotion*. In: Phillippot, P., Feldman, R. S. (Eds.), The regulation of emotion, Erlbaum, Mahwah, NJ, 2004, pp. 33-70.

[12] C.L. Stephens, I.C. Christie, and B.H. Friedman, "Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis," *Biological Psychology*, vol. 84, pp. 463-473, 2010.

[13] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J. G. Taylor, "Emotion recognition in human computer interaction," *IEEE Signal Process. Mag*., vol. 18, pp. 32-80, January 2001.

[14] J.A. Healey, "*Wearable and Automotive Systems for Affect Recognition from Physiology*," PhD thesis, MIT, Cambridge, MA, May 2000.

[15] R.W. Picard, E.Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. On Pattern Analysis and Machine Intell*., vol. 23, pp.1175-1191, October 2001.

[16] A. Haag, S. Goronzy, P. Schaich, J. Williams, "Emotion recognition using bio-sensors: First step towards an automatic system," in *Affective Dialogue Systems Tutorial and Research Workshop*, Kloster Irsee, Germany, June 2004.

[17] F. Nasoz, K. Alvarez, C.L. Lisetti, and N. Finkelstein, "Emotion recognition from physiological signals using wireless sensors for presence technologies," *Cognition, Technology and Work*, vol. 6, pp. 4-14, 2004.

[18] T. Verhoef, C. Lisetti, A. Barreto, F. Ortega, T. Zant, and F. Cnossen, "Bio-sensing for emotional characterization without word labels," Human-Computer Interaction: Ambient, ubiquitous and intelligent interaction, *13th International Conference*, San Diego, CA, USA, July 19-24, pp. 693-702, 2009.

[19] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, "Decision trees: an overviewand their use in medicine," *Journal of Medical Systems*, vol. 26, pp. 445-463, October 2002

[20] P.D. Wasserman, *Advanced Methods in Neural Computing*, New York, Van Nostrand Reinhold, 1993, pp. 35–55.

[21] K. Takahashi, "Remarks on emotion recognition from bio-potential signals," *2nd International Conference on Autonomous Robots and Agents*, Palmerston North, pp. 186-191, December, 2004.

[22] J. Arroyo-Palacios, and D.M. Romano, "Towards a standardization in the use of physiological signals for affective recognition systems," *Proceedings of Measuring Behavior 2008*, 2008.

 ISBN: 978-1-61208-250-9

Filling the User Skill Gap Using HCI Techniques to Implement Experimental Protocol on Driving Simulators

Ghasan BHATTI[1,2,3], Guillaume MILLET[1]
(1) OKTAL SA
Meudon, France
firstname.secondname@oktal.fr

Roland BREMOND[2], Fabrice VIENNE[2], Nguyen-Thong DANG[2]
(2) Paris-Est University
IFSTTAR, LEPSiS
Paris, France
firstname.secondname@ifsttar.fr

Jean-Pierre JESSEL[3]
(3) IRIT, VORTEX Group
Paul Sabatier University
Toulouse, France
firstname.secondname@irit.fr

***Abstract*—Programming activities are performed not only by programmers but also by end-users in order to support their primary goals in different domains and applications. End-users do not have formal training in programming, so interaction environment and systems are needed, which could account for user skills. The objective of our work is to fill the gap between the user skills and the goals they want to achieve using driving simulators. This paper presents the results of a research in which we have proposed a solution for the primary users of the driving simulator to design and implement experimental protocol. We have used the user-centered design (UCD) technique, conducted a user survey, and proposed a solution, in which we have categorized the Interface of the driving simulator into three sub-interfaces based on the skills of the users. These interfaces are Experiment Builder (Non-technical persons), Template builder (for technical persons) and Experiment Interface (for any user to run as experiment). A prototype based on this concept is developed and some feedback were collected from end-users. Our results indicate that, users can implement an experimental protocol without having programming skills using our proposed design.**

***Keywords*-Experimental protocol; User-Centered Design; HCI Techniques; Scenario modeling; Driving simulators**

I. INTRODUCTION

Nowadays, programs are not only written by software engineers, but also by the people (end-users) who are not expert in programming, but they do programming in order to support their primary goals. e.g. scientist write program for analysis, teachers and accountant use spreadsheets for their support, and children can use them for games and simulation. What makes these end-user programmers different from professional software developers are their goals [1].
End-users usually do not have the time to be proficient in programming so they need processes, methods and tools with immediate feedback and results, which could account for the competencies they lack to achieve their goals using programming. One of the reasons why it is difficult, that programs are abstracts [2]. It is difficult for the end-users without a programming background to think in an abstract manner to implement the specific situation compared to thinking about the same situation in the real world. So abstraction is one of the barriers, especially for novice users. It is a common observation that, the more complex the situation, the more abstraction is required in the program. Also, programming languages have not been designed by addressing the human-computer interaction issues [3]. According to John et al. [4], user interface is one of the factors besides many other factors, because it provides an interaction between the end-user and computer. So there is a need to develop a user-centered interactive environment, which could support end-users to achieve their goals.

Driving Simulator is a useful research tool for behavioral researchers to study drivers behaviors, to analyze road safety features, and for drivers training without any safety risk. They are also used to evaluate ADAS (Advance Driving Assistance Systems). ADAS are the systems to help drivers during the driving process. They are designed with the purpose of increasing the road safety. The examples of these systems include ACC (Adaptive Cruise Control), Lane Change Warning (LCW), Automatic parking etc. In order to retrieve the data (to be analyzed), researchers have to design and implement experimental protocols on their driving simulator. Designing experiments is a quite critical component in any research [5]. It requires careful consideration and controlled environment (depending on the research objectives) to achieve the desired goal of the experiment. Implementation of the experimental protocol includes the specification of subject vehicle, autonomous traffic, simulation environment, and the critical events or scenario modeling. Besides traffic and weather specification, scenario modeling is one of the steps while implementing an experimental protocol, which requires extensive programming, so end-users (behavioral researchers) need specific technical and programming skills for which they are not formally trained.

Modeling scenarios on a driving simulator is a complex and difficult task for the behavioral researchers, because of their complex and technical nature. In most of the cases, end-users are not equipped with the skills to manage these tasks. Also, in most of the existing driving simulators, the

 ISBN: 978-1-61208-250-9

scenario editors do not take into account the skills of the primary users (behavioral researchers) of driving simulator. They have to depend on the technical persons in their respective organization, who model scenarios. This is a time-consuming task for them because of dependency on the other persons. It is a time-consuming task for technical persons as well. In the case of scenario programming on driving simulators, the key challenge is to make events happen at the right time and at the right place in a non-conspicuous way, which is because of two main factors [6]. First, driving behavior is complicated and not well-understood; so it is difficult to create realistic as well as controllable traffic. Second factor is the variability of human driving behavior, as they change their speed, lane position and tactical decisions with the time during the simulation trial, which leads to variance in drivers behavior to be studied. For example, imagine you want to create an accident situation, where a car overtakes the participant car and stops in front of the participants vehicle. To implement this situation using a modeling language, there are many parameters to control and consider including speed and position of the overtaking vehicle, speed of the participant vehicle and the distance to stop from the participant vehicle. Every driving simulator provides a scripting environment, where behavioral researchers can model scenarios, by using triggers and actions or some event-driven mechanism. Besides the main issue of programming the critical situation without having enough skills, the interaction environment provided by these driving simulators does not help the user to account for the skills they lack in order to perform a task using the driving simulators. So, the objective of our work is fill the gap between the user skills and the goals they want to achieve using driving simulators .In this paper we focus on the analysis of the users, their activities and propose a solution to help them to achieve their goals while not being expert in programming.

The remainder of the paper is organized as follows. The next section describes the related work. Next, the third section introduces, our approach along with the theoretical framework Then, the fourth section provides a description of the evaluation of our approach by the end-users, which is followed by the discussion, conclusion and future work.

II. Related Work

Significant work has been done on End-User development. End-user activities take user-centered approach to develop tools and systems which can vary from, customization or parameterization, where user can set the parameters to extend or change the system functionality or presentation of the data [7] e.g. Programmorphosis [8], to the modification or the development of computing artifact from scratch e.g. Visual programming, Programming by example [9] etc.

In the rest of the section, we discuss briefly about different flavors of interaction environment for scenario modeling systems and the methodology used by the driving simulators for modeling scenarios. The purpose of our discussion is to highlight the scenario modeling approaches being used in different softwares and not to compare these softwares. We also do not discuss the details of these systems or their software components, because this is not the focus of the paper, we discuss briefly about the User Interface of some of these simulators i.e. how to place scenario objects in the scenario, and how to manipulate the scenario events.

SCANeR software [10], developed by OKTAL SA and Renault has a Graphical User Interface (GUI) for implementing the experiment protocol and to model scenarios. The scenario objects (vehicles, traffic signals..) are placed on the map using mouse. The ambient traffic is generated by placing the sources on the map or by placing individual vehicle in the map. To construct critical situations, it uses GUI for defining condition-action pair (If-else statements).

ARCHISIM [11] developed by IFSTTAR has a textual interface to define an experimental protocol and to model scenarios. The scenario objects are specified in the text files by specifying the location on the road. Ambient traffic is created manually using text files. To construct critical situations, it uses a simple text editor like notepad. In the notepad, the user can specify the condition action pair for different events.

STISIM [12] is developed by System technology, Inc. It has a textual interface for scenario modeling process. SDL (Scenario Definition Language) is a scripting language developed to define the scenario event. The scenario objects are placed in the scenario by the route traveled by the driver using SDL statements. As the participant drive in the world, the objects (buildings, traffic signals...), critical events and ambient traffic is defined using SDL.

Besides the above mentioned interfaces to model scenarios, some other simulators use different methods for scenario modeling. Wasshink et al.,[13] has proposed a movie set metaphor to generate scenarios dynamically based on Green Dino Virtual Realities Dutch Driving Simulator. They have proposed the movie set as a driving simulator, where actors (vehicles, pedestrians etc) come at the scene and play certain set of roles, which are assigned to them in the script. They have also emphasized on the problem of users to model scenarios using a scripting language. A tile-based approach is also used to specify scene and scenario elements in the driving simulator. The whole world (Terrain/Map where participant drive the car) is divided into different tiles, which are configured, assembled and then loaded into the driving simulator during the experimental trial. A Tile is a section of the route on which contains the elements like roads, traffic signals, buildings, trees and scenario objects. These tiles are then grouped together and loaded using an interface or by specifying the sequence of the tiles. The scenario events are then created on the map. In some systems, tiles are static and may not be altered or moved during the experiment run [14], while there are some systems in which tiles as well as data on the tiles (Scene objects, scenario objects) can be altered dynamically during the experiment run [15]. Sometimes it is required to generate a script dynamically during the simulation. Some driving simulators provide this functionality, and some use other means to fulfill this requirement. All these systems and methods have not been designed and developed by keeping in mind the skills of the primary users (behavioral researchers) of the driving simulators. Also, interaction environment of these systems is not so user-friendly that they could account for user skills. So there is a need for a user-centered interaction environment, which could enable users to program scenarios without having extensive programming skills.

III. METHOD

In this section, we present the methodology that we have followed to achieve our goals. As we are following the UCD approach. we have involved the users at the very start, and conducted a survey of primary users of driving simulators. After that, we have proposed and designed the solution and took initial feedback from the users.

A. User Survey

A user survey was conducted to get to know about the users problems and ideas to model scenarios on driving simulators. We interviewed 19 driving simulator users with various backgrounds using different simulators. Most of the users we interviewed were using ARCHISIM and SCANeR softwares. The focus of the interviews was to discuss about their problem, needs and requirement while modeling scenarios on driving simulator. For detailed methodology and results see [16]. In this survey, users have explained their problems and have given ideas about the interface of driving simulator which could fill the gap between their skills and the goal they want to achieve, regardless of any software.

1) Main problems identified:

1) Controlling the ambient traffic around the participant during the critical event.
2) Tuning or optimizing the critical events.
3) Finding relevant functions in the scripting language to perform a task.
4) Optimizing and debugging the script.
5) Selection of triggers (time or distance) to model a critical event.

2) Main ideas proposed:

1) Drag and drop the critical events/situations using the mouse. These critical events should be editable at low-level using the low-level scripting language of the driving simulator.
2) Interaction of the users with the map. Users want to interact with the map while modeling the scenarios.
3) Preview of the activity they are performing to design a complete experimental protocol.

In our survey, 7 out of 19 users had no programming experience at all, and they had never programmed scenarios by themselves while implementing an experimental protocol. 9 out of 19 users had very little programming knowledge. So they have to take total or partial help from the technical persons while modeling scenarios. The users had an average 2 years experience of working on driving simulators.

B. Proposed Approach

Every existing scenario modeling system provides a mean to interact with the scenario modeling software, which is textual or GUI. But these interfaces still do not fill the gap of user skills and goals. Based on the discussions with end-users during the survey, we identified the following steps which end-users undertake while designing and implementing an experimental protocol.

1) Terrain selection
2) Configure participant/subject (Initial position, ADAS, speed and other platform dependent parameters).
3) Configure the ambient traffic (Number of vehicles per hour, their configuration i.e. speed, itinerary...).
4) Configure the environment (weather condition, light..).
5) Set the dependent variables to be collected.
6) Construct critical events using scripting.
7) Experiment execution and data collection (variables).

We propose a new User Interface (UI) and experiment development approach based on the problems raised and the ideas proposed by the users during the survey and the aforementioned steps. Traditionally, in order to implement an experimental protocol, a user (technical person or researcher) uses the same interface for implementing an experiment protocol, regardless of level of their technical and programming skills.

In the proposed design we split the scenario modeling activity into 3 sub-interfaces depending on the roles they have to perform while modeling scenarios and the set of skills they have. The 3 roles ”Researcher” as ”R1” (Low or no programming skills), ”Technical Person” as ”R2” (Good programming/technical skills) and Operator as ”R3” (No technical skills required) correspond to the Experiment Builder, Template Builder and Experiment Interface, respectively.

We explain this new approach and interface with the help of an example. Our example scenario contains two events.

- **Accident Event:** A vehicle overtakes the participant vehicle by increasing its speed and changes its lane to the lane of the participant vehicle and then brakes.
- **Pedestrian cross event:** A pedestrian walks and then crosses the road as the participant vehicle approaches.

1) Template Builder: This sub-interface will be used by technical persons performing role R1 having good programming and technical skills. R1 will design the GUI-based templates of the scenario events. The template builder

 ISBN: 978-1-61208-250-9

will let R1 use existing functions offered by the scenario-modeling environments of the driving simulator to model scenario events. In our example, at the back-end, for the template "Accident", the "Template Builder" will let R1 program a vehicle around the participant vehicle to accelerate, change position, and brake at some distance from the subject vehicle. For the event "Pedestrian crossing", the "Template Builder" will let R1 program a pedestrian to walk and cross the road as the participant vehicle approaches the intersection. At the front end of the template, there would be different text fields to specify the parameters for the events "Accident" and "Pedestrian crossing". These parameters will be filled by the Researcher in the "Experiment Builder" sub-interface, if he or she wants to modify the default values. The templates developed by R2 will be stored in a template library, so that researchers could access the templates in the 'Experiment Builder' interface as described in Figure 1.

2) Experiment Builder: This sub-interface will be used by researchers/trainers performing role R2 and possibly have low or no programming skills. R2 will define the whole experiment using a user-friendly and intuitive GUI which includes specifying experiment condition, environment, ambient traffic, and data to be collected. To specify the critical events or situations to be studied, R2 will access the template library developed by technical persons in the 'Template Builder' and will place them in the scenario editor using drag and drop and proceeded by a user-defined trigger. If needed, R2 will fill the parameters of the template.
In our example, end-user will specify the position and actors involved in the templates "Accident" and "Pedestrian crossing" besides the template parameters, if needed. There will always be default appropriate values for all the parameters of the template.

3) Experiment Interface: The experiment interface will be used by user (researcher or the person who will execute the experiment, depending on the organization culture) performing role R3. Using this sub-interface, R3 will load and execute the scenarios in the driving simulator, developed with the "Experiment Builder". R3 can change the parameters of the scenario or template (if needed), during the experiment trial and finally will collect the data to be studied. We will be focusing more on sub-interface "Experiment builder", because researchers are our focused users, who do not usually have programming skills. So in the next section we will be talk about our prototype building experience of "Experiment Builder".

C. Prototype development

A prototype based on this concept is developed. In this paper, we are focusing on the Experiment builder sub-interface, which will be used by researchers. We do not discuss the solution in detail because of the space constraints, but we try to present the transformation of user requirements into feature to fill the skills gap of the researchers, which is

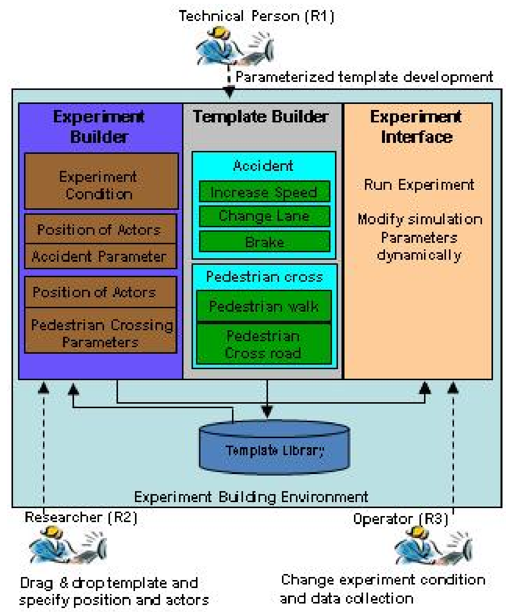


Figure 1. Scenario Modeling Process

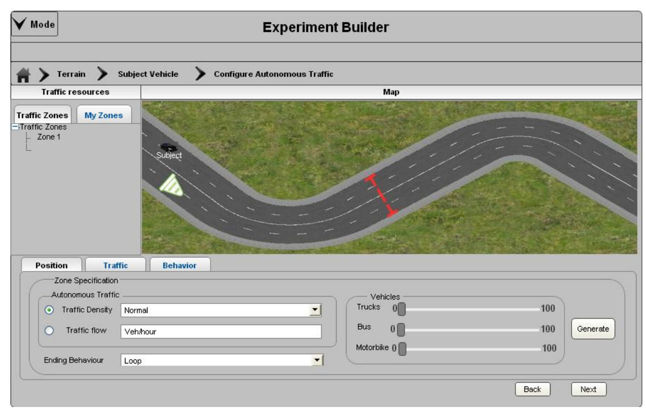


Figure 2. R2 specifying Autonomous traffic

the focus of this article. The prototype is based on the problems identified, user suggestions and the steps that are identified from the user survey. The prototype is built using "Justinmind Protyper" [17], a tool to develop rich interactive wire frames. It is difficult to explain all the steps because of space constraints, but we explain the main steps proposing solution to user problems. The user is guided using the Breadcrumb navigation [18] during the experiment building process, as you can see in Figures 2, 3 and 4. Targeting problem 1 in User survey section, the end-user can specify the autonomous traffic at higher level by specifying and configuring the traffic zone, rather than scripting individual vehicle as specified in the Figure 2. Targeting problems 2, 3, 4 and 5, a user can create critical events as specified in the Figure 3.The user will just drag and drop the template in the scripting area and customize the template. By customization,

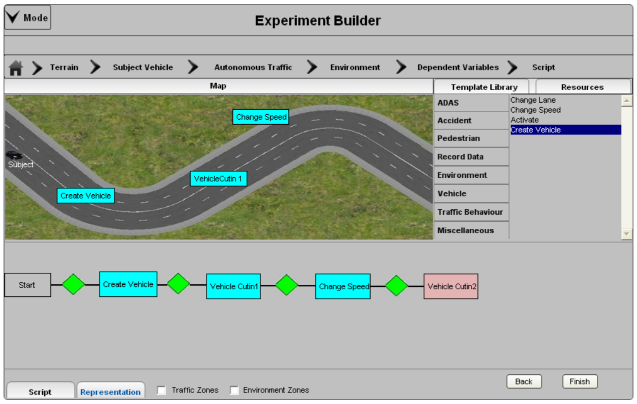


Figure 3. Top: 2D representation of map and resources. Bottom: R2 Speciying critical events using drag and drop

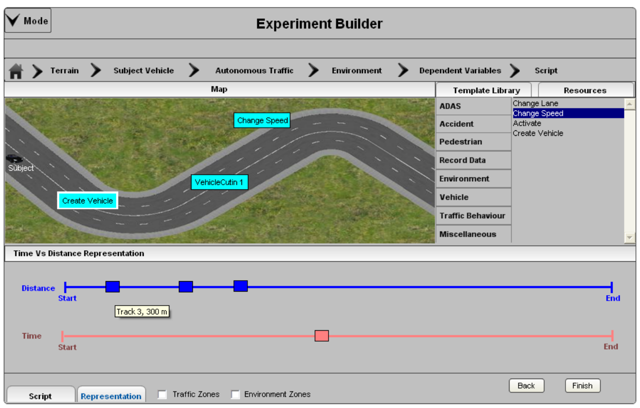


Figure 4. Top: 2D representation of map and resources. Bottom: Temporal and Spatial representation of the experiment

users can easily optimize and tune the critical events. We have also tried to incorporate the user ideas of interacting with the map, i.e. during the creation of environment zone in the environment step, during the creation of traffic zones in the autonomous traffic step and at the last step, where the user creates the critical events. So users can drag the templates from the template library and drop them on the map or in the scripting area. The light blue templates on the map indicate that they are configured based on position of the participant vehicle, and the pink one's indicate that they are configured based on simulation time (See Figure 3).

The division of experimental protocol in different steps enables user to create the experiment in the way they imagine, because traditionally, they have to perform all the steps of experimental protocol at one place, which is fuzzy to tackle, if the scenario is long and complex. In order to keep the scripting process transparent for the users, the whole experiment can be viewed as temporal/spatial representation as specified in the Figure 4.

D. Prototype Evaluation and Results

Feedback was collected from users about our prototype. It was qualitative evaluation, as it was the initial phase of the design, and we were interested to discuss the issues in detail for a subjective evaluation. Nine users were involved in this evaluation. These users were the primary users of the driving simulators who design and implement the experimental protocols on driving simulators and have low or no programming skills. The approach was explained to the user and then they performed a small exercise on the prototype. Users were observed during the exercise and interviewed later on. The interviews were not recorded and average time for a user was 30-40 minutes with 5-7 minutes for explanation, 15-20 minutes for the exercise and 10-20 minutes for the interviews. In the interviews, users were asked about their feedback with respect to the new approach, division of experiment protocol steps, the difficulties during the exercise i.e. which step was difficult to comprehend, and they were also asked that, is there anything that needs to be improved. The questions were not limited to what we have specified or prepared. We discussed in details, if users had raised any issue. In the exercise, the users created a small scenario, in which first, they had to create a lead vehicle to follow, which was followed by a critical situation in which they had to enable a vehicle to cut-in the participant vehicle. After this event, there was another event, in which the vehicle being followed brakes. And in the last there was another cut-in situation as shown in the Figure 3.

The structure of UI and the steps to create the experimental protocol were quite clear to all the users. They all gave a quite positive feedback about the approach for creating scenarios/events by using templates. There were some minor problems, for example, 4 users told that creating a zone for autonomous traffic was not very intuitive and a bit uncomfortable, and that was observed for 2 more users. Three users also suggested that, in order to create events based on position, they should be able to drop the template on the map. Before they had to drop the template in the scripting area and later they were able to change the position of the template by dragging it on the map. Five users attempted to do that as well. During the interviews, they felt easy and intuitive to configure the templates, rather than doing the low-level coding, and they were also interacting with the map, so they were aware of the simulation activity that where and what is going to happen.

IV. Discussion and Conclusion

We have discussed in detail about the user-centered design to develop experimental protocols on driving simulators. In this paper, we have focused on one class of end-users (behavioral researchers), who are the primary users of the driving simulators. The objective is to fill the gap between user skills and the goals they want to achieve in an effective way using driving simulators. We have separated the role

 ISBN: 978-1-61208-250-9

of technical person, who will implement the core logic of the critical events in an abstract way in the form of event templates using low-level language provided by the driving simulator. In this way, end-users will not have to go through a typical programming process. All they have to do is just to drag and drop the templates from the template library and configure them. It will reduce their time in programming and also the dependency on the technical persons. If the researchers have enough skills and want to develop complex situations by themselves, they can change their role into technical persons and use the Template Builder sub-interface to edit a template or create a new template by themselves. During the user survey, users specified that, making some changes in an existing scenario would be easier for them than writing the complete scenario from scratch.

Another aspect of our solution is the division of the implementation task of experimental protocol procedure into meaningful subtasks. In existing systems, users have to use low-level languages to configure each step in the scripting environment. So it is difficult to tackle complex and long scenarios, even if some users have good programming skills. We have split the whole experimental protocol development procedure into different steps, and have provided support to the users during each step by keeping in mind the skills of the users, which make our approach different from the existing approaches.

Now, we have evaluated the concept and design at its initial phase, We also plan to conduct a controlled experiment to evaluate the solution in detail. After the user evaluation, we have improved the user-interface, and now we are implementing the complete system. As driving simulators vary based on their execution platform, so we need to generalize this solution for different platforms, and we are also working on that at the moment.

Acknowledgment

This work has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement no. 238833/ ADAPTATION project. www.adaptation-itn.eu

References

[1] A. J. Ko, R. Abraham, L. Beckwith, A. Blackwell, M. Burnett, M. Erwig, C. Scaffidi, J. Lawrance, H. Lieberman, B. Myers, M. B. Rosson, G. Rothermel, M. Shaw, and S. Wiedenbeck, "The state of the art in end-user software engineering," *ACM Comput. Surv.*, vol. 43, no. 3, pp. 21:1–21:44, April 2011.

[2] A. Cypher and D. C. Smith, "Kidsim: end user programming of simulations," in *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 27–34.

[3] A. Newell and S. K. Card, "The prospects for psychological science in human-computer interaction," *Hum.-Comput. Interact.*, vol. 1, no. 3, pp. 209–242, Sep. 1985.

[4] J. F. Pane, B. A. Myers, and L. B. Miller, "Using hci techniques to design a more usable programming system," in *Proceedings of the IEEE 2002 Symposia on Human Centric Computing Languages and Environments (HCC'02)*, Washington, DC, USA: IEEE Computer Society, 2002, pp. 198–.

[5] R. A. Romano and H. A. Stoner, "A quick and effective prototype experimental design and analysis tool," in *Driving Simulation Conference North America*, Iowa, IA, 2007.

[6] K. Joseph and G. Timofey, "Scenario authoring," in *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*. CRC Press, Apr 2011.

[7] H. Lieberman, F. Patern, M. Klann, and V. Wulf, "End-user development: An emerging paradigm end user development," in *End-User Development*, ser. Human Computer Interaction Series. Springer Netherlands, 2006, vol. 9, pp. 1–8.

[8] A. Ioannidou, "Programmorphosis: a knowledge-based approach to end-user programming," in *INTERACT*, 2003.

[9] A. Cypher, "Eager: programming repetitive tasks by example," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '91. New York, NY, USA: ACM, 1991, pp. 33–39.

[10] G. Reymond, A. Heidet, M. Canry, and A. Kemeny, "Validation of Renault's dynamic simulatorfor Adaptive Cruise Control experiments," in *Driving Simulation Conference DSC*, Sep. 2000.

[11] S. Espie, F. Saad, B. Schnetzler, and F. Burlier, "Microscopic traffic simulation and driver behavior modeling: the archisim project." in *Strategic Highway Research Program and Traffic Safety on Two continents*, Lille, France, 1994.

[12] G. D. Park, R. W. Allen, and J. T., Rosenthal, "Flexible and real-time scenario building for experimental driving simulation studies," in *CHI 2011*, Palo Alto, CA, 2011.

[13] I. Wassink, B. van Dijk, J. Zwiers, A. Nijholt, J. Kuipers, and A. Brugman, "In the truman show: Generating dynamic scenarios in a driving simulator," *IEEE Intelligent Systems*, vol. 21, no. 5, pp. 28–32, Sep. 2006.

[14] Y. Papelis, O. Ahmad, and G. Watson, "Developing scenarios to determine effects of driver performance: Techniques for authoring and lessons learned," in *Driving Simulation Conference North America*, Dearborn, Michigan, 2003.

[15] P. Suresh and R. R. Mourant, "A tile manager for deploying scenarios in virtual driving environments," in *Driving Simulation Conference North America*, Orlando, Florida, 2005.

[16] G. Bhatti, R. Bremond, J.-P. Jessel, G. Millet, and F. Vienne, "User requirements to model scenarios on driving simulators," in *5th International conference on Driver behaviour and Training (ICDBT)*, Paris, France, 2011.

[17] "Justinmind prototyper." [Online]. Available: http://www.justinmind.com/ (Accessed: 29/01/2013)

[18] J. Nielsen, "Breadcrumb navigation increasingly useful," in *Jakob Nielsen's Alterbox*, April 10, 2007. [Online]. Available: http://www.nngroup.com/articles/breadcrumb-navigation-useful/ (Accessed: 29/01/2013)

Instrumentation and Features Selection Using a Realistic Car Simulator in Order to Perform Efficient Single-User Drunkenness Analysis

Audrey ROBINEL
LAMIA,
Université Antilles-Guyane
Pointe-à-Pitre, France
audrey.robinel@univ-ag.fr

Didier PUZENAT
LAMIA,
Université Antilles-Guyane
Pointe-à-Pitre, France
didier.puzenat@univ-ag.fr

***Abstract*—We instrumented a car simulator by gathering low level data and fed it to an artificial neural network in order to perform blood alcohol content (BAC) estimations. The results depend on the quality of the data extraction and processing, and also on the selected inputs. We explain our data extraction and processing methodology, and how we used it to generate reliable and comparable features. At last, we describe the performances of individual features and how they combine. In the end, the prototype was able to accurately estimate the BAC value of a subject after being trained with driving samples of this subject with various BAC values.**

***Keywords*-Blood Alcohol Content; Driving; Interface; Artificial Neural Networks; Intelligent systems; Machine learning; Instrumentation; Interface; Regression; Classification; Feature Selection; Car; Simulator;**

I. INTRODUCTION

A. Problematic

Drunk driving is a major cause of accidents on the road. Alcohol alters people behaviour and ability to drive properly in many ways that are well known. We are working on a way to determine whether an user is in condition for driving or not, and estimate this user's blood alcohol content. The most common device for such a task is the breathalyzer. We however proposed a method for doing so without using a breathalyzer or any invasive device, by only monitoring how the user performs in ordinary tasks (the general methodology was described in [1]). In the current case, we monitor a car simulator controls and any available data from the embedded calculators, and analyze the driver's behaviour (we first presented the method for ”car” monitoring in [2] with a racing game, and later with a car simulator in [3]). In order to do so, we had to train an artificial neural network (ANN) in order to make it capable of patterns detection within monitored data in real time. In this paper, we will discuss the methodology for data extraction and processing in order to perform this task. We will also discuss the issues that can impair the results, and present the monitored data. We will then describe some of the obtained features, along with results showing the performances of the network for each feature.

Figure 1. Our realistic car simulator (left) and its “force feedback” steering wheel with pedals (right).

B. Methodology summary

We will now provide a quick summary of the methodology (more on this in [1], [2] and [3]).
Our subject did consume an amount of alcohol, and then drove in our realistic car simulator (“Stars AF 2011”, presented in [3] and provided by “ApportMédia”) on a highway scenario with a countryside road at the end (Fig. 2). The simulator runs on a personal computer equipped with a Logitech G25 force feedback steering wheel with pedals (depicted in Fig. 1)

While the subject was driving, numerous low level data related to the controls and other parameters have been monitored. Before driving, the blood alcohol content value had been measured with a consumer class breathalyzer. We thereby obtained a set of monitored data associated with a BAC value. The subject repeated the operation (we call that a run) multiple times in order to provide examples with various BAC values so that we could create an example base.

Our final goal was to estimate the BAC value of the driver by feeding the monitored data to an ANN (a classical multilayer perceptron with back-propagation learning based

 ISBN: 978-1-61208-250-9

on the FANN library [5]). Whereas we could monitor many low level data, not all were suitable for our task. Furthermore, among the compatible data, we had to make some processing in order to be able to feed it to the ANN. We will at first describe how we gathered the low level data from the simulator. Thereafter, we will discuss how it had to be processed to produce features (potential ANN inputs), and afterwards present some of the obtained features. At last, we will present some results and discuss the impact of features selection on the ANN results.

C. On the realistic aspect of the simulator

When we use the term "realistic", we don't mean that the prototype used could replace a real car. Visually, we consider it realistic because it represents what an user would see in a car. It is not photo-realistic, but the cars are modelled after real cars, and so are the roads and environment. The software is considered realistic enough to be used for learning drivers in driving schools in France and Europe. However, in no mean we expect neither our prototype nor the software to replace real cars experimentation. Indeed, the user sees the virtual world through a screen that lacks the peripheral vision accessible in a car, and in this version, the controls are different from a real car controls (this point will howbeit be addressed in the next experiment, since we will be using real car controls in a hardware simulator with the same software). Furthermore, motion sensations are lacking, and various details too. Nevertheless, our prototype is realistic enough to analyze driving behaviour alterations caused by alcohol, and to validate our methodology and hypothesis.

Figure 2. The simulator software features "realistic" graphics.

II. Instrumentation of the simulator and monitoring of the controls.

A. low level data collection

The first task was to monitor low level data by instrumenting the simulator. Our simulator uses a script language that enabled us to read data from the physics engine (acceleration, speed ...) of the simulator, the position of objects (distance between cars, position on the road ...), and the state of all user controls (steering wheel, pedals, etc.). However, we can only read the instant value of those data at a given time. In order to be able to analyze the driver's behaviour, we have to be able to work on the evolution of the values of the variables. Since we had no way of knowing what data may or may not be useful before running the ANN, we decided to gather all the accessible values as often as possible. The simulator refreshes the values 50 times per second (one computation cycle per frame displayed on screen, and the frame rate is*fixed* at 50 per second), and all our variable can be read at each frame. We thereby obtain a vector of 28 values each 20ms. Gathering data at a constant time interval is important, since it enables us to compute the variation in time of each value easily.

B. collected data

An exhaustive list of the 28 variables instrumented is presented in Table I. The first column indicates the id of the variable, and the second gives a description of the value monitored. Although most are self-explanatory, we will detail those who are not. The "range" column indicates the interval of values than can be expected, completed by the type in the next column to describe the ensemble of possible values.

The variables 22 to 25 describe the surface under each wheel. In our scenario, it can be either 1 for road, or 2 for off-road. Other values are used for snow and other environments that was not present in our scenario.

We also gather the position of all vehicles, including the user's, but it is unused for now (other than for computing the distance to the closest vehicle in variable 14). Obviously, not all measures are meaningful in each experiment : here we decided to use an automatic gearbox car in the simulator, which explains why we don't use the clutch pedal readings.

C. Creation of suitable inputs from low level data

When the subject has finished driving on a scenario, we obtain a matrix of data that must be processed in order to be usable by the ANN. As we are not using a time based ANN model, we must generate suitable potential inputs (features) for the ANN. The simplest way to obtain a feature that reflect the behaviour of the subject on an interval of time is to use a statistical indicator, such as the average of the value. For some variable, such as vehicle speed, it provides a suitable feature whereas for other data (such as variable 20), it would not provide meaningful or comparable data . Therefore, we must for each variable study how it can be exploited and generate the corresponding features.

D. On normalization of features

Since the duration of the monitored run can vary, we had to ensure that the features used would be comparable. On the one hand, some features give the number of occurrences of an event, and must then be divided by the duration of the run. This provides us the frequency of the considered event, which does not depend on time (of course, in order to remain a significant parameter, the duration has to be long enough). On the other hand, other features were generated using averaged values, or similar computation. In this case, there is no need to divide the feature by time, as they are

Table I
LOW LEVEL MEASURES INSTRUMENTED FROM THE SIMULATOR

ID	Name	Range	Type	Notes
0	Accelerator pedal position	0;1	float	
1	Brake pedal position	0;1	float	
2	Clutch pedal position	0;1	float	unused in this experiment
3	Steering wheel position	-1;-1	float	
4	Vehicle speed	0;60	float	
5	Vehicle roll	-1;1	float	
6	Vehicle pitch	-1;1	float	
7	Vehicle engine RPM	0-8000	int	unused.
8	Instantaneous fuel consumption	0+	float	unused.
9	Instantaneous CO_2 production	0+	float	unused.
10	Instantaneous CO production	0+	float	unused.
11	Instantaneous HC production	0+	float	unused.
12	Instantaneous PM production	0+	float	unused.
13	Instantaneous NO_X production	0+	float	unused.
14	Distance to closest vehicle	0+	float	
15	Front wheel sleep angle	-1;1	float	
16	Rear wheels sleep angle	-1;1	float	
17	Vehicle acceleration on $\vec{X}$ axis	-1;1	float	
18	Vehicle acceleration on $\vec{Y}$ axis	-1;1	float	
19	Vehicle acceleration on $\vec{Z}$ axis	-1;1	float	
20	Engine gear ratio	{0,1,2,3,4,5,6}	int	0=no gear engaged, and 6=back. Unused.
21	Lights state	{0,1,2,3}	int	unused
22	Surface under front left wheel	{0,1,2,3,4,5}	int	1=road, 3=offroad
23	Surface under front right wheel	{0,1,2,3,4,5}	int	1=road, 3=offroad
24	Surface under rear left wheel	{0,1,2,3,4,5}	int	1=road, 3=offroad
25	Surface under rear right wheel	{0,1,2,3,4,5}	int	1=road, 3=offroad
26	Left indicator state	0 or 1	bool	unused
27	Right indicator state	0 or 1	bool	unused

already comparable. When needed, we used time as our divider. But we also could have used the number of measures used for the considered feature. Some features can have quite different ranges, so in the end, all values should be normalized relatively to each other to use the same range (in our case 0 to 1 or -1 to 1). We did so, and we often noted an improvement of experimental results when doing it.

E. How much pre-processing should be done on the data?

We made a design choice of not introducing human intelligence in the analysis of data (or as few as possible). We could indeed probably have simplified the problem for the ANN by constructing more complex, and higher level features that would be very specific to the problem of blood alcohol content. It may thus reduce the amount of work necessary for ANN optimization. However, doing so would also make our device dedicated to our current problem, whereas we wanted it to be usable for a wider class of problems related to drivers. We still have to provide results on other problems, but the designed prototype will be usable with no further modification.

Overall, we tried to create generic and and simple features in order avoid being specific.

III. PRECISION OF THE MEASURES

Precise measurement is important in order to create an example base. Indeed, the more precise are the measures, the less loss of accuracy we can expect on the trained ANN.

A. On instrumented variables

For the instrumented variables, in our case, we had no problem with the accuracy of the variables instrumented, as we use a simulated environment. The measures was therefore completely exact whereas when instrumenting a real device, the accuracy of the measurements must be taken into account. However, if we can have access to numerous variables, it becomes possible to select the most accurate for features generation. Furthermore, the combination of multiple features may compensate inaccurate measures.

B. Breathalyzer accuracy issues

The other part of the problem is the measurement of the BAC of the subject (or the expected output of the network, in a general case). We conducted tests on several subjects, and found that our breathalyzer can provide noisy measures, as shown in Fig. 3. In the first case, the BAC value increases of $+0.11g.l{-}1$ in 7 minutes before and after a decrease. In the second case, it decreases by $-0.09g.l{-}1$ in 5 minutes in the second case. The decrease of BAC value in the second case is 6 times bigger than the average alcohol elimination value of human beings, and in the first case, the variation is quite unstable. We can conclude that the variations shown here depicts an important noise in measures, thus degrading our device accuracy. In the first case such an increase is not a problem, however we should have a constant variation, and not an increase preceded and followed by decreases, considering that the subject did not consume any alcohol for 20 minutes. For this subject, we think that an important part of the variation is noise. This is not the case for all subjects, but in this experiment, we could not have more reliable measures with a consumer class device. In the worst cases presented here, the noise could be up to $0.1g.l^{-1}$, which degrades the accuracy that we can obtain.

Ideally, we should have used a law enforcement class breathalyzer, but the costs for such a device is considerably higher than ours. Furthermore, obtaining such a device may be difficult due to the fact that they are not meant to be sold to regular citizens.

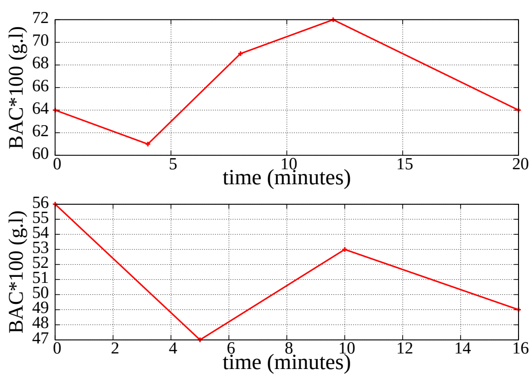


Figure 3. Evolution of BAC over time for two subjects.

When excluding some subjects that had excessive variations, we increased the accuracy of the prototype. However, we decided to keep as many subjects as possible in order to have more generalist results.

IV. Obtained features

After processing the low level data collected, we generated 25 features, that are presented in Table II.

For each feature obtained, we have an identifier (F0 to F24), and a descriptive name. Most are self explanatory, but we will describe those who are not. The third row indicates if this feature is already normalized or not. Then the "from" row tells what low level data from the Table I was used to generate the feature.

For the following rows, we present results that will be detailed later on. The idea is that we created a training base using only one feature, trained the network, and then tested it in generalization. We present regression results (R_x) that will be detailed in section V.

As mentioned earlier, we did not use the gear controls in this experiment. We however plan to use the clutch pedal and gearbox selector in future experiments. It must be noted that we can generate other features from the car's controls, but we decided to start with a few and then expand the list. We can thus consider the use of sub-features such as the average variation of a pedal when increasing the pressure, and when decreasing the pressure, or the duration of an action on the controls (average brake use duration, average duration of acceleration pedal increase time, decrease time...)

For now, we will explain some the used features. Many are related to the steering wheel, and steering wheel actions. We define a steering wheel action as a sequence of measures that starts when the steering wheel leaves the neutral (centre) position, and ends when it comes back to this neutral position. F0 gives the average duration of those actions. F6 is a variation in which we consider only the part of the action where the user turns the wheel in a constant direction. We stop the timer when he reverts the rotation of the wheel. F5 gives the count of direction changes operated on the steering wheel. F3 monitors how often the user was turning to one side before moving the steering wheel in the opposite side (e.g. the wheel was turned left, then reverts to neutral before going into the right half). For F1 and F4, we use "proximity alerts", which means that the distance to the closest vehicle goes below a fixed value. F1 gives the amount of measures where it occurred, while F4 counts the number of sequences where it happened (a sequence begins when the subject drives below the safety distance and ends when he goes farther. The set of measures is counted as *one* proximity alert sequence).

V. Results

A. Measurement of the ANN success rate in generalization

Using K-Fold cross validation, we test our network in generalization. When we perform classification tasks, if the output of the ANN corresponds to the expected class, we count a valid response, and otherwise we count an error. In the end, we divide the number of valid responses by the number of examples tested in generalization.

For regression purposes, we had to introduce a maximal tolerated error, ϵ. For each value returned by the ANN, we compute the distance between this value (N) and the expected value (E) : $dist = |E - N|$. If it is below a fixed epsilon, we count a success. Otherwise, we count a failure. We then compute the success rate of the ANN in generalization.

B. Analysis of the individual features results

The subject performed a total of 28 runs, which corresponds to approximately one hour and a half of driving in the simulator. Due to the fact that most of the examples had

Table II
FEATURES EXTRACTED FROM LOW LEVEL MEASURES.

The "source" row gives the corresponding measure used from table I. R1 gives individual success rates for $\epsilon = 0.2$, and R2 for $\epsilon = 0.1$. Both are expressed in percent. "Avg error" gives the average error of the network in $g.l^{-1}$.

ID	Name	Normalized	Source	R1:$\epsilon = 0.2$	R2:$\epsilon = 0.1$	Avg error
F0	Average steering wheel action duration	yes	3	75	50	0.126
F1	# of proximity alert measures	no	14	92	46	0.124
F2	# of steering wheel actions	no	3	92	53	0.108
F3	# of car direction changes	no	3	78	50	0.106
F4	# of proximity alert sequences	no	14	75	46	0.106
F5	# of changes of rotation direction of steering wheel	no	3	89	50	0.122
F6	Avg duration of constant direction steering wheel actions	yes	3	78	50	0.124
F7	# of measures with a wheel out of the road	no	22-25	78	53	0.108
F8	# of measures with any wheel out of the road	no	22-25	85	53	0.117
F9	Average vehicle speed	yes	4	85	53	0.118
F10	Average steering wheel shift from neutral pos.	yes	3	96	46	0.109
F11	Average roll of the car	yes	5	85	53	0.111
F12	Average pitch of the car	yes	6	82	50	0.112
F13	Front wheels avg sleep angle	yes	15	85	50	0.118
F14	Rear wheels avg sleep angle	yes	16	82	46	0.123
F15	Average accelerator pedal position	yes	0	85	46	0.119
F16	Average brake pedal position	yes	1	82	46	0.119
F17	Average steering wheel position	yes	2	82	42	0.113
F18	Average vehicle acceleration on $\vec{X}$ axis	yes	17	96	53	0.116
F19	Average vehicle acceleration on $\vec{Y}$ axis	yes	18	96	53	0.111
F20	Average vehicle acceleration on $\vec{Z}$ axis	yes	19	82	50	0.120
F21	Average accelerator pedal variation	yes	0	85	46	0.116
F22	Average brake pedal variation	yes	1	78	46	0.120
F23	Average steering wheel variation	yes	3	100	50	0.108
F24	Average vehicle speed variation	yes	4	50	50	0.121

BAC values below $0.5g.l^{-1}$, we could not create an unbiased classification base, and then only present regression results. In the R1 column, we present the features individual success rates for $\epsilon = 0.2$ and in R2 for $\epsilon = 0.1$. When looking at the results of R1 ($\epsilon = 0.2$), we note important variations, with a results ranging from 50% for F24 to 100% for F23. Some features seem to be much more discriminant than other. As one feature reaches up to 100% of success rate, there is no point in combining inputs. We therefore decreased ϵ to 0.1. In that case, individual results are much lower(42 to 53%), and little variation exists between most features. It is hard do discriminate the best features, and we can see that the best features in R1 are not always the best features in R2.

C. Combining features

Now we will combine features in order to see how the network performs. When using the 4 best features, according to the results of R1 (F10,F18,F19,F23), we obtain a 75% success rate, with an average error of 0.089764 for $\epsilon = 0.1$ (do note that R1 results are for $\epsilon = 0.2$). With only one feature, we obtained at best 53%, and combining features improved significantly the results.

We will now combine features according to R2. However, the results of the features are close, so we can have multiple configuration of 4 of the best features. When using F2,F7,F8,F9, we reach a 71.43% success rate and an average error of 0.091848. With another configuration, F2,F7,F11,F18, we reach 78.57% and an error of 0.080105. In both cases, we kept an ϵ of 0.1.

When we use the 4 best features according to the average error (F2, F3, F4, F23), we obtain a 78.57% success rate, and an average error of 0.077460. This configuration reaches the best success rate until now, and obtains the lowest average error.

We tried to combine various features, without considering if they are among the "best" or not, but rather by using features related to varied controls or data. A good combination was F2 F7 F13 and F18, with a 82% success rate and an average error of 0.085317. Using various sources provided to offer good results, so we tried with more inputs (F2 F8 F13 F18 F19 F22) and reached 85.71% with an average error of 0.088704.

VI. CONCLUSIONS

Our prototype showed that it could reliably estimate the BAC of a subject. We were able to obtain success rates up to 85% in generalization, when training the device with less than one hour and a half of driving. If integrated in a real car, it means that we could quickly gather data to create a learning base. Of course, for blood alcohol content estimation, there would be some practical complications for the creation of the base. For all that, the described methodology can be applied to many physiological parameters estimation.

When a simple and cheap solution is available, it may not be interesting to use our method. However, some physiological parameters of a subject are quite complex to acquire, requiring invasive, pricy or long procedures. In such cases, our method could prove to be quite useful : the cost of the device in both time and money only has to be spent once to create a large enough example base, and won't require invasive procedures when used in real life situation. Furthermore, on contrary to some methods that provides a measure with a delay, we can provide estimations in a short time, and continuously, so that the variation of the monitored parameters could be considered.

The main downside of this method is that it may require a long search for the most suitable features. The quality of the features used must also be ensured : whereas the ANN won't use bad features, including those makes the selection of good inputs longer. Creating the system can be quite complex, but it should provide a reliable, transparent, fast, non-invasive and economic way of estimating user parameters. Of course, the more precise are the measures, the more accurate the prototype should be. This implies that increasing the initial investment in both cost of used measurement hardware and number of examples monitored should provide an improved device while maintaining exploitation costs at constant level.

The selection of features can't be done by only using the individual success rates of features for a fixed ϵ. As a matter of fact, the best features changed from a ϵ to another. When combining the best features according to the individual average error, we obtained the lowest average error, but not the best success rate. It seems hardly feasible to obtain a simple metric to establish the performance of a feature, considering that individual scores may not reflect the potential effectiveness of combined features. A more complex solution could be to use something like the saliency of Optimal Brain Damage [6] : after having trained the network with all features, we could study how the suppression of an input cell corresponding to a feature impacts the network's performances. This value could provide a good metric.

VII. Perspectives

In this paper, we presented the results of a single user example base, but we will try to create multi-user bases in order to determine if the system could learn from multiple subjects so that it could estimate the BAC of an unknown subject. Although it was not the scope of this paper, it will be interesting to see how multi-user bases impacts the results presented here.

Our next goal will be to proceed with experimentation on tiredness and attention using the same software, but in an hardware simulator, with realistic car controls. Instrumenting will be done on the same basis, but with more available data (such as gear ratio, pedal...). The use of the simulator hardware should provide a driving experience closer to real cars, and enable us to collect more accurate data. We are also considering the use of data related to events rather than the average behaviour of the subject, like variation of various parameters when specific events occur (e.g. an accident, a dangerous situation, a change of the driving conditions...).

In the long term, we are looking forward to conduct similar experiments into real cars or trucks. However, we need to keep improving our prototype in the simulator in order reduce the number of experimentation needed in real cars, so that we can reduce the potential cost of development.

We are also looking forward to determine whether or not a subject is able to drive, and if not, what impairs his skills up to the point that he or she should not drive. This will be much more complex and require a collaboration with researchers in medical science. It should however be a good illustration of a complex to establish diagnostic about the subject that can't yet be done automatically.

Overall, we want to be able to detect various causes that alters the ability to drive of a subject.

Acknowledgment

This work has been funded by ApportMédia (www.apportmedia.fr), *la Région Guadeloupe* (www.cr-guadeloupe.fr), and *the European Social Fund* (ec.europa.eu/esf).

References

[1] D. Puzenat and I. Verlut, Behaviour analysis through games using artificial neural networks. In *proceedings of the Third International Conferences on Advances in Computer-Human Interactions* (ACHI 2010), pages 134-138, Sint Maarten (Netherlands).

[2] A. Robinel and D. Puzenat, Real time drunkenness analysis through games using artificial neural networks. In *proceedings of the Fourth International Conferences on Advances in Computer-Human Interactions* (ACHI 2011), pages 206-211, Gosier (France).

[3] A. Robinel and D. Puzenat, Real Time Drunkenness Analysis in a Realistic Car Simulation. In *ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0 pages 85-90. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420. (ESANN2012)

[4] ApportMédia, "StarsAF2011" realistic car simulator product online documentation www.simucar.com/logiciels/STARS_AF.html.

[5] S. Nissen, Implementation of a Fast Artificial Neural Network library (FANN). Technical report, Department of Computer Science University of Copenhagen, October 2003.

[6] LeCun and Al., Optimal Brain Damage. In Advances in Neural Information Processing Systems (NIPS 1989), vol 2

Adaptive Simulation of Monitoring Behavior: The Adaptive Information Expectancy Model

Bertram Wortelen, Andreas Lüdtke
Human Centered Design Group
OFFIS - Institute for Information Technology
Escherweg 2, 26121 Oldenburg, Germany
Email: {wortelen|luedtke}@offis.de

***Abstract*—Human attention is a fundamental but limited resource. Especially when performing safety critical tasks a suitable distribution of attention is essential for safe operation. E.g., changes in task relevant information have to be recognized in time in order to react adequately. This paper presents the Adaptive Information Expectancy (AIE) model, which simulates the scheduling of attention within cognitive architectures. It can be used for model-based evaluations of interactive human-machine systems. Results of a first evaluation study are shown based on a simple laboratory monitoring task. An overview on the AIE model is given and it is shown how it was integrated in the Cognitive Architecture for Safety Critical Task Simulation (CASCaS). A formal model for the laboratory task was developed and then simulated using CASCaS. Several aspects of the AIE model are evaluated on the basis of the simulations of this agent in two main steps. The first step of the evaluation compares the agent behavior with results from the studies conducted by Senders. In this step two alternative AIE model variants are compared to participants' behavior. The second evaluation step explores parameter sensitivity and the convergence behavior of the model.**

***Keywords*-Event expectancy; cognitive model; attention allocation; monitoring behavior;**

I. Introduction

Detailed knowledge about human attention allocation is vital for designers of human-machine interaction, e.g., in cars or aircrafts. It has been acknowledged by many researchers that executable cognitive models have the potential to capture such knowledge and to make it readily available to designers (e.g., [1][2][3]). This paper presents the Adaptive Information Expectancy (AIE) model which is an extension of the seminal SEEV model introduced by Wickens et al. [4].

How humans distribute their attention depends on several factors. The SEEV model is a predictive model of attention distribution that relates the amount of attention allocated to a specific information source to four influencing factors. The abbreviations of these factors form the acronym SEEV: Saliency of information events, Effort required to perceive the information, Expectancy of new information events and Value of the task, that requires the information. The SEEV model can easily be applied to estimate percentage dwell times (PDTs) – the percentage of time a human operator spends looking at an information source, e.g., of a human-computer interface.

The AIE model is based on the two knowledge driven factors Expectancy and Value. Although it does not consider Saliency and Effort, it extends the SEEV model in two ways. (1) It relates the attention distribution to an executable task model, which can be simulated, e.g., in a cognitive architecture. Based on the simulation further measures like gaze frequencies or link values can be estimated, besides the mere prediction of PDTs as provided by the SEEV model. (2) The second extension is related to the operationalization of Expectancy. The SEEV model requires a system designer or Human Factors Expert (HFE) who applies the model to give an estimate of each of the influencing factors for all considered information sources. For this, Wickens et al. [4] propose a lowest ordinal algorithm as an easy to use method, that orders the influencing factors by small integers according to their rank. Although this method has been proven simple and effective, it is only a very rough operationalization that is dependent on the subjective rating of the HFE. The AIE model strives to replace this by deriving the expectancy factor dynamically from a simulation of the task model in a dynamic environment. It is thus able to adapt its attention distribution automatically to the current situation, which is an enhancement over the SEEV model. It furthermore provides a much more detailed view because it integrates the simulation of task performance and the simulation of attention control in a tightly coupled way. The long term goal of this research is to use the AIE model to predict the allocation of attention dependent on design characteristics of human-machine interfaces and associated tasks in complex and safety critical environments.

The following shows results of a preparatory study that was used to evaluate the AIE model on the basis of a laboratory monitoring task, which was developed by John Senders in the 1960s [5]. The paper starts with an overview on the AIE model and its integration in the Cognitive Architecture for Safety Critical Task Simulation (CASCaS) [6][7]. Then an overview on Senders' task and the derived task formalization is given in Section III and IV. Section V is dedicated to the detailed evaluation of the AIE model.

 ISBN: 978-1-61208-250-9

II. Adaptive Information Expectancy Model

The AIE model and its integration into CASCaS will be described here briefly on an abstract level to provide a basic understanding. For a more elaborated description see [7].

Cognitive architectures can be understood as engines that are intended to execute formal models of tasks in a psychological plausible way. CASCaS is a modular architecture consisting of components for perceptual, memory, knowledge processing and motor processes. The AIE model is a general model of attention allocation and thus is integrated in CASCaS as part of a general model of human cognition. In contrast, task models describe task specific aspects of human behavior. The interpretation of a task model by the cognitive architecture eventually simulates human-like behavior. In the following the term *(cognitive) agent* refers to the combination of task model and cognitive architecture.

A. Control of Attention

While the SEEV model is typically used to predict the visual attention allocation to multiple information sources, the AIE model intends to predict to which task an agent mentally attends. However there is typically a very strong relationship between visual and mental attention, which Just and Carpenter [8] named the eye-mind-assumptions. This assumption was adopted for the AIE model evaluation presented in Section V, where the gaze behavior of the participants in Senders' study is compared with the gaze behavior of the cognitive agent.

An assumption of the AIE model is, that human behavior is goal oriented and every task serves to achieve one specific goal. The AIE model is applicable to situations where multiple tasks have to be performed in parallel in a time-shared fashion, like e.g., a pilot that has to monitor a set of displays, while controlling the aircraft and communicating with the pilot non flying. Attention is a limited resource and often only one of the tasks can be processed consciously. Although in real situations it is often possible to execute some parts of tasks really in parallel, this aspect will not be discussed in this paper.

To instantiate a cognitive agent CASCaS loads a hierarchical task model. Each task seek to achieve a goal and is modeled by a set of rules. These rules represent the knowledge of the human operator about the task. For a driver model, for example, they describe how the driver interacts with the car and the surrounding traffic. The rule language is based on the well-known GOMS notation [9]. All rules consist of a left-hand side (IF) and a right-hand side (THEN). The left-hand side names the goal to be achieved and a Boolean condition that defines in which situations the rule shall be applied. The right-hand side defines the actions that are executed when applying the rule.

Multitasking situations in the task model have been handled in the past by a very simple mechanism that treats every task as equal and repetitively executes every task for a certain but short amount of time in a fixed sequence. This mechanism is now replaced by the AIE model. The AIE model assigns a weight to each goal g_i of all active tasks:

$$w(g_i) = U \cdot \frac{u_{g_i}}{\sum\limits_{g_j \in G} u_{g_j}} + V \cdot \frac{v_{g_i}}{\sum\limits_{g_j \in G} v_{g_j}}. \qquad (1)$$

In the above equation G is the set of goals for all active tasks, u_{g_i} is the expectancy coefficient that describes how much the agent expects new information for the task of goal g_i and v_{g_i} is the value or importance of goal g_i. Thus the weight $w(g_i)$ depends on the relative importance of a task compared to the importance of all tasks and the relative information expectancy of a task compared to all tasks. The factors U and V are used to adjust the overall influence of task importance and information expectancy.

CASCaS now selects the next goal to be executed in a probabilistic way. The probability of selecting goal g_i is defined by the relation of all weights:

$$P(g_i) = \frac{w(g_i)}{\sum\limits_{g_j \in G} w(g_j)} \qquad (2)$$

In Equation 1 the value and expectancy factors are linked by addition. However in applications of the SEEV model additive combinations (e.g., [10]) as well as multiplicative ones (e.g., [11]) can be found. Arguments can be found for both variants. This was discussed by Wickens et al. [12]. They achieved a better model fit with the additive formulation. But still there is no consensus about this issue. To shed further light on this matter, the AIE model was implemented in both variants. To explicitly distinguish between both variants the symbol AIE$^+$ is used for the additive formulation and AIE* for the multiplicative formulation.

B. Event functions

To use the AIE model the coefficients in equation 1 have to be defined for each task. For the value coefficients v_{g_i} it is suggested to employ the lowest ordinal algorithm that is typically used for the coefficients of the SEEV model [4]. But for the expectancy coefficients u_{g_i} an automatic operationalization is proposed.

Wickens et al. describe the expectancy factor as an *"information-related measure of event expectancy (e.g., bandwidth, event rate; [...])"* [4, p.3]. This view is adopted by the AIE model. The expectancy coefficients are operationalized on the basis of information events. An event is defined as follows:

If at time t information, which is relevant for a goal g is used to achieve g, then $e = (g, t)$ is called an event at time t for goal g. Let E be the set of all events that occurred during a simulation, then $E_g \subseteq E$ is denoted to be the set of all events for goal g. Events can be ordered and indexed by their time of arrival. With $e_{g,i}$ the i-th event in E_g is denoted.

The identification of events in CASCaS is straight forward. CASCaS processes tasks using a rule engine similar to other rule based architectures like e.g., ACT-R. It provides four different rule types: regular, percept, waiting and reactive rules. When a task gets actively processed by a rule engine the process is typically as follows. If the task needs some information appropriate percept rules are executed that instruct CASCaS to direct its gaze to an information source, which provides the information. After the information has been perceived, it can be used by regular rules to achieve the task goal. If the perceived information is of no use for the agent, a waiting rule is fired, which signalizes, that another task should be executed. Hence the execution of a regular rule corresponds to the recognition of an information event for the goal, which is supported by that rule.

During the simulation of the task model CASCaS records the events of all goals and develops for each a cumulative frequency distribution function H_g of the distances between consecutive events $e_{g,i}$ and $e_{g,i+1}$. The value $H_g(\Delta t)$ can answer the question, how often an event occurred not later than Δt time units after the previous event. This is used to describe whether the agent can expect new events for a specific goal and thus the expectancy coefficients are defined by $u_g = H_g(t - t_{g,n})/d_g$, with t being the current time, n being the index of the last event for g and d_g being the amount of time that the agent was working on g. Thus u_g is a time dependent function which is called the event function of g. One effect of this operationalization is that the expectancy of new events continuously increases since the last event was observed.

Another effect is that the behavior of the agent changes over time. At the beginning of the simulation it has no knowledge about event distance. But the more events the agent detects the more stable the event functions get. Thus the behavior should change less the more time passes. The learning speed of the agent should correspond to the convergence speed of the event functions. According to the Berry-Esseen theorem [13] the pointwise convergence speed should be bounded by $\mathcal{O}(n^{-\frac{1}{2}})$, with n being the number of events which are recorded in H_g. If the distribution of event distances does not change over time, having a fixed average event rate, this can be expressed in a time-dependent way by $\mathcal{O}(t^{-\frac{1}{2}})$. Unfortunately the assumption that the distributions do not change over time is false, because there is always a feedback within the cognitive agent when using the AIE model: The event functions influence the selection of goals; the selection of goals determine, where the model looks at; where the model looks at influences the perception of events distances; finally, the perception of event distances influence the event functions. But assuming, that the effect of this feedback loop is small, at least a similar learning speed should be obtained. This was considered during model evaluation and will be addressed in Section V.

Table I
BANDWIDTHS OF THE GAUGES FOR THE THREE EXPERIMENT CONFIGURATIONS.

Configuration	Participants	Signal bandwidths per gauge (Hz.) 1	2	3	4	5	6
C1	5	0.08	0.16	0.32	0.64	-	-
C2	3	0.03	0.05	0.12	0.20	0.32	0.48
C3	2	0.02	0.04	0.08	0.16	0.32	0.64

III. SENDERS' MONITORING TASK

For a first evaluation of the AIE model a simple laboratory task was selected. It is the monitoring task developed by Senders [5]. A cognitive agent was developed that relies on the AIE model and is able to interact with this task. The model is evaluated in section V against data that Senders obtained in his studies on this task. In this section a short overview on the task and the setup he used in his studies is given. For more details see the original work [5].

In the Senders Task participants had to observe a set of gauges that displayed dynamic values of currents for fictitious devices. Every time one of the displayed signals fell below -45 μA or above 45 μA participants had to push a button. Senders investigated how bandwidths of the signals influence the gaze distribution of the participants. He used five different tasks configurations. Three of these are used for the evaluation and are denoted by C1, C2 and C3. The signal bandwidths of each configuration are listed in Table I. It must be said, that the configuration C1 belongs to a different study than configurations C2 and C3. It was conducted before the study involving C2 and C3.

C1 investigates the gaze behavior while monitoring four gauges. Five participants executed this task for 1 h per day over 30 days . Gaze behavior of the last three minutes of each day have been analyzed.

C2 was similar to C1, but six instead of four gauges were used. Participants executed the task for 1 h per day over 10 days. Gaze behavior of the last 11 minutes of the last day have been analyzed. This configuration is shown in Figure 1, which shows the geometrical layout of the six gauges (left side) and the viewing distance of the observer (right side).

C3 was similar to C2, but signal generation was changed, which resulted in a different set of signal bandwidths.

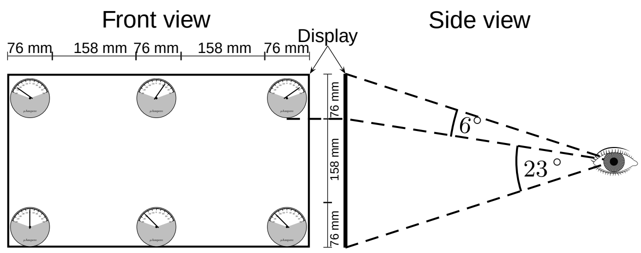


Figure 1. Task configuration with six gauges (reconstructed from [5]).

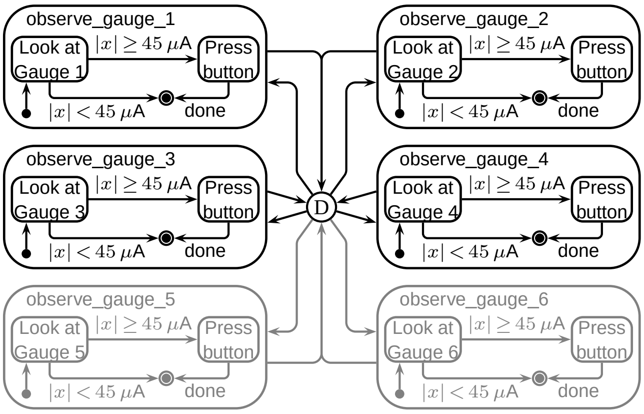


Figure 2. State diagram of the cognitive agent's task model. The states observe_gauge_5 and observe_gauge_6 are only active for C2 and C3.

IV. Cognitive Agent

Senders Task was modeled using the rule based language of CASCaS. Figure 2 shows the semantics of these rules in the form of a state chart. The overall goal is to observe all gauges. This is decomposed into one subgoal for the observation of each gauge. This results in four subgoals for C1 and six subgoals for C2 and C3. These subgoals are represented in Figure 2 by the top-level states. The task that is executed to achieve each goal is depicted within the top-level states by the process steps that have been mentioned in Section II-B. At the beginning of the task processing the agent executes a percept rule, which instructs CASCaS to look at the information source that provides information for the task. This is in this case the gauge for the specific task. If the perceived information demands a reaction of the agent a regular rule is fired. For Senders' task it happens when the signal is in the alarm region ($|x| \geq 45\,\mu$A). The response button is pressed as reaction by the execution of a regular rule. If the signal is not in the alarm region no reaction is required, the agent fires a waiting rule and the task is finished at least for the moment. In the condition that the signal is in the alarm region, the agent uses the perceived information to trigger an action by executing a regular rule. Thus these situations are the events for this task.

The AIE model comes into play at the decision point marked with a *D* in the figure. Here the agent selects which subgoal it will process next. According to the AIE model the agent will select to observe a gauge, where it highly expects an alarm, in order to detect as much alarms as possible. The expectancy coefficients u_i will be derived during the simulation. The task values v_i have to be assigned by the model developer. Because all tasks have the same priority, these coefficients have been selected to be 1 for each task. Expectancy and Value factors are weighted equally ($U = V = 1$).

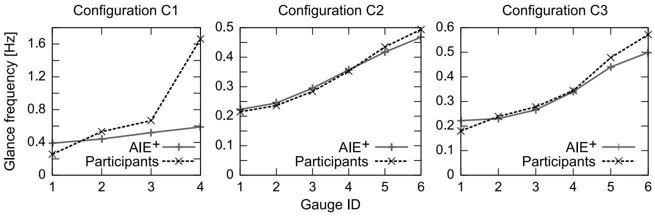


Figure 3. Glance frequencies of participants and AIE$^+$ agent for each gauge in each configuration.

V. Evaluation

To evaluate the AIE model CASCaS was connected to a simulation of the three configurations of Senders Task. The cognitive agent described in the last section was instantiated in CASCaS. CASCaS itself was connected to a simulation of the three configurations of Senders Task.

A. Additive Variant: AIE$^+$

The evaluation starts with an analysis about the model fit of the AIE$^+$ model to Senders' data. The agent was simulated 10 times for 3 h of simulation time in each of the three configurations. To avoid that the results are affected by learning effects, only the last 1 h from every 3 h simulation run was analyzed.

In [5] Senders presented the glance frequencies of the participants to each gauge. In Figure 3 this data is shown together with the glance frequencies that have been observed in the simulation of the AIE$^+$ agent. The figure shows, that the agent's behavior well matches the experimental data for C2 and C3 with very high trend correlation of $R^2 = 0.996$ for C2 and $R^2 = 0.984$ for C3. Also the absolute deviations measured by the root-mean-square error (RMSE) are small with RMSE = 0.04 Hz for C2 and RMSE = 0.09 Hz for C3. The situation is different for C1. Although the agent shows the general trend ($R^2 = 0.851$), the absolute deviation is quite high (RMSE = 1.1 Hz). Especially the overall frequency for C1 is with 3.1 Hz considerably greater than for C2 (2.0 Hz) and C3 (2.1 Hz). The cognitive agent is not able to reach the high glance frequencies in C1, because the model that is part of CASCaS and calculates the duration of saccades and fixations does not permit such small gaze durations, which are required to simulate these high gaze frequencies. For details on this model see [14]. The reason for the high frequencies in C1 compared to C2 and C3 are not known. It might be due to the additional 20 days of practice that participants had for C1, or just due to differences in the eye data processing, which was done by a frame-by-frame rating for videos of participants' eye movements.

However, Senders also calculated link values for C1. The link value probability according to ISO 15007-1:2002 describes for two information sources the relative frequency of gaze transitions between these information sources compared to all observed gaze transitions. The link values

 ISBN: 978-1-61208-250-9

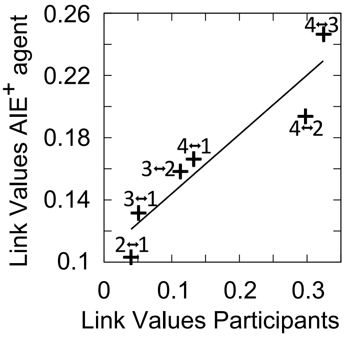


Figure 4. Correlation of link value probabilities for configuration C1. A data point i↔j represents the link value probability between gauge i and gauge j.

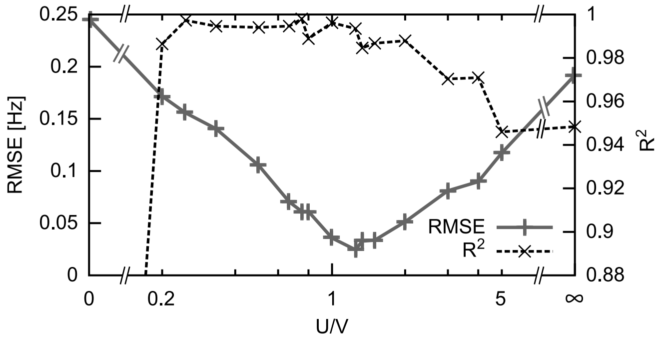


Figure 5. RMSE and R^2 values obtained from simulating the AIE$^+$ agent with different values of U/V.

obtained during the simulation of the AIE$^+$ agent correlate well with the ones observed by Senders for C1 (R^2 = 0.87) as can be seen in Figure 4.

B. Sensitivity Analysis

In equation 1 it can be seen, that there are quite a lot of coefficients that have to be determined before using the model. Compared to the SEEV model the AIE model already eliminates the need to define the expectancy coefficients u_i by expert knowledge.

But still the value coefficients v_i and the general weights U and V remain. A high number of parameters bears the danger that it allows, in principal, to fit the model predictions to any kind of data. Wickens addresses this issue by proposing the lowest ordinal algorithms to restrict the choice of coefficient for the application of the SEEV model. This approach is also proposed for the value coefficients of the AIE model. However, for the monitoring task discussed in this paper this is meaningless, because Senders did not instruct the participants to prioritize any of the gauges. Therefore the value coefficients are all equal and thus are not free parameters at least for this agent.

In applications of the SEEV model the issue of weighting the influence factors Expectancy and Value differently is typically not addressed and the factors are weighted equally. The same was done for the weights of the AIE$^+$ agent presented in the previous section ($U = V = 1$), which led to very good results for C2 and C3. Nevertheless these are free parameters, which have only the weak restriction that usually they are chosen equally. Although it seems that U and V are two parameters, it is effectively one parameter. Inserting equation 1 into equation 2 leads to:

$$P(g_i) = \frac{\frac{U}{V} \cdot \frac{u_{g_i}}{\sum\limits_{g_j \in G} u_{g_j}} + \frac{v_{g_i}}{\sum\limits_{g_j \in G} v_{g_j}}}{\frac{U}{V} + 1} \qquad (3)$$

Note that this conversion is only valid for $V \neq 0$. As can be seen now U and V only occur as U/V and thus this is only one free parameter. A sensitivity analysis was conducted to estimate the effect of the relation U/V on the model fit. A set of selected relations were analyzed:

0/1, 1/5, 1/4, 1/3, 1/2, 2/3, 3/4, 4/5, 1/1, 5/4, 4/3, 3/2, 2/1, 3/1, 4/1, 5/1, ∞

For each relation 10 simulations with a duration of 3 h have been executed like described in Section V-A. The obtained RMSE and R^2 values are displayed in Figure 5. The highest correlation values have been obtained in the range from 0.2 to 2.0 with $R^2 > 0.98$, while the lowest absolute deviations have been found in the range between 1.0 and 1.5 with a RMSE < 0.04 Hz. So the popular choice of $U = V = 1$ was obviously also for this task an adequate one, although a slightly better fit has been observed with a slightly higher value of U. This is a satisfactory result. According to Pitt et al. [15] a model should be stable around a reasonable region of parameters. As an equal weighting of expectancy and value is a usual assumption, this property seems to be well met by the cognitive agent.

It should be noted that the considerations made in this section are only valid for the presented task model. For a more general view on the AIE model this work has to be repeated for a set of different task models in different application domains.

C. Multiplicative Variant: AIE*

In Section II it was mentioned, that there is no consensus about how expectancy and value are linked. In the following the simulation results are shown using the AIE* agent. In the multiplicative variant of equation 1 the weighting factors U an V are eliminated when the fraction in equation 2 is reduced. Thus there is no free parameter. Executing the simulation again in the way described in Section V-A using the AIE* formulation produces the glance frequencies shown in Figure 6.

It can be seen, that the model fit is worse than for the AIE$^+$ agent. The differences are clearly visible for C2 and C3 for which the AIE$^+$ agent showed a very good model fit. Here the AIE* agent especially underestimates the frequencies to the gauges with low signal bandwidths.

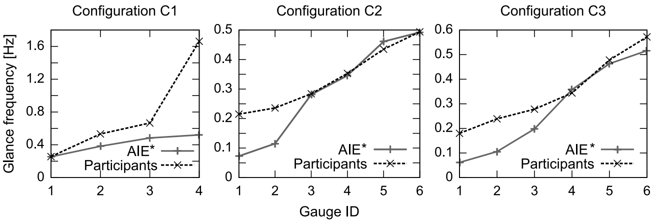


Figure 6. Glance frequencies of participants and AIE* agent for each gauge in each configuration.

Giving the data a closer look reveals that the obtained frequencies are almost identical to the frequencies of the AIE^+ agent using weights $U = 1$ and $V = 0$. And both models are in fact identical. The value coefficients of the AIE^* model disappear when reducing equation 2, because they are all equal. This is equivalent to disabling the value factors in the AIE^+ model ($V = 0$).

D. Learning Convergence

In Section II-B it was assumed that the learning speed should be bound by $\mathcal{O}(t^{-\frac{1}{2}})$ if feedback effects can be neglected. To investigate this the same agent as in Section V-A was used. But now 20 simulation for C2, each with a duration of 9 h were made. Every 5 minutes the event functions of all 20 simulations were pairwise compared using the difference measure V from Kuiper's test, which describes the similarity of two frequency distributions u_1 and u_2. The symbol V^k is used to avoid a mix-up with the weight of the value factors. According to [16] V^k is calculated by:

$$V^k = \max_{0<\Delta t<\infty}(u_1(\Delta t) - u_2(\Delta t)) + \max_{0<\Delta t<\infty}(u_2(\Delta t) - u_1(\Delta t)) \quad (4)$$

This resulted in 190 comparisons every 5 minutes. It was expected that the event functions are getting more and more similar over time and thus the average V^k should converge

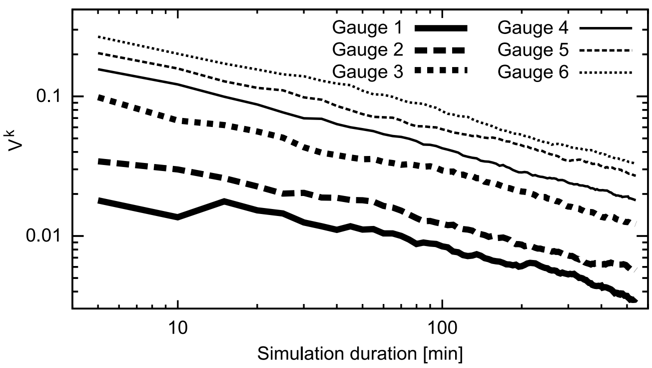
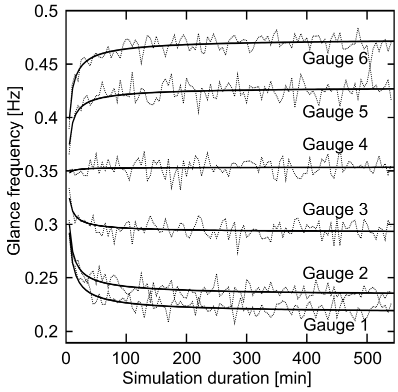


Figure 7. Convergence of event functions. Differences between event functions of different simulation runs measured by V^k are displayed on a log-log plot and asymptotically approach 0.



Gauge	a	c
1	-0.082	0.216
2	-0.072	0.233
3	-0.035	0.292
4	0.006	0.354
5	0.058	0.429
6	0.081	0.475
b	0.682	

Figure 8. Learning effect on glance frequencies. Frequencies over time are fitted to functions of the form $a \cdot t^b + c$. Fitted parameters are listed on the right side.

towards 0. The average V^k values are plotted on a log-log graph in Figure 7. It can be seen, that these values form straight lines for the event functions of each task. This strongly supports the initial assumption, that the learning speed can be expressed by a function of the form $a \cdot t^b$. It also supports the assumption that the event functions always converge against the same function.

In the same way the consequences of the learning process on the glance frequencies of the cognitive agent were analyzed. It turned out, that these reflect the learning process. They develop according to functions of the form $a \cdot t^b + c$. This knowledge now allows to fit the simulation data to such functions, and to estimate their asymptotical value. In Figure 8 this can be seen for C2. In the left graph the average glance frequencies are plotted over time by dotted lines. With solid lines functions of the form $a \cdot t^b + c$ are plotted that were fitted to the data using the method of least squares. In the table on the right side the fitted coefficient are listed. The c-coefficient is the estimation of the asymptotical glance frequency value. This graph helps to identify how much simulation time should be dedicated to learning the event distance distributions. For the AIE^+ agent a learning phase of at least 30-50 minutes should be used. After this time there is only little change in the glance frequencies.

The same analysis was conducted for the AIE^* agent. It required a much longer learning time. The parameter b that determines the learning speed for the glance frequencies is for the AIE^* agent only at $b = 0.432$.

VI. Conclusion and Future Work

The AIE model supports the simulation of task models within the cognitive architecture CASCaS, by providing a model of attention control. It was shown how the AIE model automatically derives expectancy for information events and uses this to guide its attention. A good model fit was achieved between the agents glance frequencies and results taken from studies conducted by Senders [5]. The issue of combining the Expectancy and Value factors additively or multiplicatively was addressed by evaluating both variants. The additive variant provided a better model fit. A sensitivity analysis for the free parameter revealed that the agents behavior is stable within a reasonable parameter region. It was analyzed how fast the agent is able to recognize the distribution of events. The hypothesized speed function $a \cdot t^b$ was successfully fitted against the observed simulation data.

The AIE model is strongly related to the SEEV model and extends it in some ways. However, it is more an alternative for the SEEV model than a replacement. The SEEV model provides a simple and fast way to estimate the distribution of attention. But its static way of application does not provide a detailed insight in the situation under investigation. This is provided by the AIE model, as it is integrated in a cognitive architecture as basic simulation framework. The AIE model additionally benefits from the simulation, because it dynamically derives expectancy values during simulation, which in contrast has to be done by human factors experts for the SEEV model. Thus from an application point of view the AIE model trades off simplicity for richness of detail.

It should although be noted, that the results presented in this paper are only valid for the investigated monitoring task and the transferability as well as scalability remains to be investigated. There are some aspects that reduce the representativeness of the study. Senders did not manipulate the information value for the gauges. Thus changes in the value coefficients are not addressed. However the expectancy coefficients are the main focus of the AIE model. In addition, the participants did not influence the displayed signals, which is unrealistic for most human-machine systems. This application served as a first evaluation step for the model. More and richer applications are required to ground the findings of this study. A subsequent step to this study is the application of the AIE model to a cognitive car driver model. In the long term, this work shall lead to a general evaluation method for human machine interaction that is based on virtual human-in-the-loop simulation.

ACKNOWLEDGMENT

Parts of this research has been conducted within the European project D3CoS and is funded by ARTEMIS-JU and national authorities under grant agreement no. 269336.

REFERENCES

[1] B. E. John, S. M. Starr, and B. S. Utesch, "Experiences with collaborative, distributed predictive human performance modeling," in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2012, pp. 437–452.

[2] J. A. B. D. Dzaack, "Analyse kognitiver benutzermodelle fr die evaluation von mensch-maschine-systemen," Ph.D. dissertation, Technische Universitt Berlin, Fakultt fr Verkehrs- und Maschinensysteme, 2008.

[3] F. E. Ritter, G. D. Baxter, G. Jones, and R. M. Young, "User interface evaluation: How cognitive models can help," in *Human-computer interaction in the new millennium*, J. M. Carroll, Ed. Addison-Wesley, 2001, pp. 125–147.

[4] C. D. Wickens, J. Helleberg, J. Goh, X. Xu, and W. J. Horrey, "Pilot task management: Testing an attentional expected value model of visual scanning," University of Illinois, Aviation Research Lab, Savoy, IL, Tech. Rep. ARL-01-14/NASA-01-7, November 2001.

[5] J. W. Senders, "Visual scanning processes," Ph.D. dissertation, University of Tilburg, Netherlands, 1983, lawrence Erlbaum Assoc., Hillsdale, NJ,1984.

[6] A. Lüdtke, J.-P. Osterloh, L. Weber, and B. Wortelen, "Modeling pilot and driver behavior for human error simulation," in *Digital Human Modeling*, ser. Lecture Notes in Computer Science, V. G. Duffy, Ed. Springer, Berlin, 2009, vol. 5620/2009, pp. 403–412.

[7] B. Wortelen, M. Baumann, and A. Lüdtke, "Dynamic simulation and prediction of drivers' attention distribution," unpublished.

[8] M. A. Just and P. A. Carpenter, "A theory of reading: From eye fixations to comprehension," *Psychol. Rev.*, vol. 87, no. 4, pp. 329–354, July 1980.

[9] S. K. Card, T. P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*. Hillsdale: Erlbaum, 1983.

[10] B. F. Gore, B. L. Hooey, C. D. Wickens, and S. Scott-Nash, "A computational implementation of a human attention guiding mechanism in MIDAS v5," in *Digital Human Modeling*, ser. Lecture Notes in Computer Science, V. G. Duffy, Ed. Springer, Berlin, 2009, vol. 5620/2009, pp. 237–246.

[11] W. J. Horrey, C. D. Wickens, and K. P. Consalus, "Modeling drivers visual attention allocation while interacting with in-vehicle technologies," *J. Exp. Psychol.-Appl.*, vol. 12, no. 2, pp. 67–78, 2006.

[12] C. D. Wickens, J. S. McCarley, A. L. Alexander, L. C. Thomas, M. Ambinder, and S. Zheng, "Attention-situation awareness (A-SA) model of pilot error," in *Human performance modeling in aviation*, D. C. Foyle and B. L. Hooey, Eds. New York: CRC Press/Taylor & Francis Group, 2008, pp. 213–239.

[13] W. Feller, *An Introduction to Probability Theory and its Applications*. John Wiley & Sons, 1966, vol. 2, ch. 16.

[14] J.-P. Osterloh and A. Lüdtke, "Analyzing the ergonomics of aircraft cockpits using cognitive models," in *Proceedings of the 2nd International Conference on Applied Human Factors and Ergonomics (AHFE)*, W. Karwowski and G. Salvendy, Eds., Las Vegas, Nevada, USA, July 2008.

[15] M. A. Pitt, W. Kim, D. J. Navarro, and J. I. Myung, "Global model analysis by parameter space partitioning," *Psychological Review*, vol. 13, no. 1, pp. 57–83, 2006.

[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.

 ISBN: 978-1-61208-250-9

Shape Modeling: From Linear Anthropometry to Surface Model

Ameersing Luximon
Institute of Textiles and Clothing
The Hong Kong Polytechnic University
Hong Kong
tcshyam@polyu.edu.hk

Huang Chao
Institute of Textiles and Clothing
The Hong Kong Polytechnic University
Hong Kong
huang.chao@connect.polyu.hk

***Abstract*— Due to technological innovations, there is a shift from linear anthropometric measures to surface model. This is essential since body surface information is needed in medical, archaeological, forensic and many other disciplines. As a result, there is a need to shift from linear anthropometry tables to surface anthropometry databases. This study provides a general modelling technique, to convert linear anthropometry to complex surface anthropometry using recursive regression equations technique (RRET) and scaling technique. In order to build the model, scanned data of a 3D body part for a selected population is needed. After scanning, the parts are aligned along a reference axis before extracting cross-sections perpendicular to the reference axis. Using the cross sections, a 'standard' part is computed based on an averaging method. The 'standard' part provides the shape information of a given population and thus it can be stored in a database for shape prediction. Using the recursive regression equations technique, regression equations are constructed based on selected anthropometric measures of each cross section. During 3D part shape prediction, only few anthropometric measures are used and the anthropometric measures of all the other cross sections are predicted recursively using the equations developed from RRET. Eventually, the predicted measures are used to scale the 'standard' part in order to generate a predicted 3D shape of the selected part. Previous studies published by our team on foot shape modelling have given accurate results of about 2mm. Thus, depending on different applications, this technique can be applied to generate 3D shape from anthropometry and can be applied to reconstructive surgery, forensic, anthropology, design, psychology and other fields.**

Keywords-anthropometric; surface antropometry; recursive regression equation.

I. INTRODUCTION

According to the World Health Organization [1], anthropometry, a method to assess the size, proportions and compositions of the human body using simple equipments, is universally applicable, inexpensive and non-invasive. Anthropometry, even though dating back to Renaissance [2], emerged in the nineteenth century largely by German investigators in the physical anthropology discipline, while they needed to study the quantitative description of the human body reliably [3, 4]. The basic anthropometric techniques were developed during that time and they are still used today [3]. The different anthropometric measures are represented in percentile values in anthropometric tables [2, 5] and since the values are statistical values, they cannot be combined to create a single human body [6]. The anthropometric percentage values are generally used to compare different populations and to design for a given population. Anthropometric data has been widely used in fields ranging from engineering to arts. It has been used in equipment and workplace design [6, 7], forensic investigation [8, 9], growth and nutrition evaluation [4, 10], medicine [11-13], archaeology and cultural studies [14-16], and sports science and fitness evaluation [17-19]. Anthropometric studies have been widely carried out because of its non-invasiveness [1], inexpensiveness [4], simplicity [1], portability [20] and reliability [21]. Furthermore, as the anthropometric dimensions vary among different groups of population, anthropometric tables have been developed based on age, race, region, and occupation [2]. Even though, there are many studies on anthropometry, most of the tables have only linear anthropometric data. Since surface geometry is required in many applications, more research is required to enhance the existing data on linear anthropometry.

Surface anthropometry describe the size and shape as well as the 3D surface geometry of the human body [22]. It is possible to combine surface scan data and internal measurements [23], thus anthropometric techniques can be used to find the size, shape and proportion of the external as well as internal structures of the body. Thus in this modern world, data collected from Magnetic Resonance Imaging (MRI), Computed Tomography (CT) or Computed Aided Tomography (CAT), sound, optical (laser or structured light) or any scanning devices can be used to create surface anthropometry of the external as well as the internal structures of the body. Since surface anthropometry provides information on the complex surfaces of the human parts, in addition to the common linear anthropometry measures, it can be used for many applications such as planning and assessment of facial surgery, design and manufacture of implants and prostheses, facial reconstruction in forensic applications, archaeology, psychology, genetics, and

 ISBN: 978-1-61208-250-9

comparative and evolutionary anatomy [24]. In addition, there are growing uses of synthesize 3D digital animated images of human models in science fiction movies and 3D digital dummies for equipment testing [25]. Although surface anthropometry seems to be very useful, it has several disadvantages. Surface scanning equipment is relatively expensive and is not widely available. It is relatively difficult to operate and require special skilled technicians to capture the dynamic and complex body shape. Furthermore, the data obtained from the scanning equipment requires additional processing and statistical data analyses are not trivial resulting in only few large-scale studies. Still, surface anthropometry is very useful for different applications and there is a need to simplify the method to acquire surface anthropometry, reduce the cost of equipments and develop surface anthropometry databases. In this study, a method to predict accurately the surface anthropometry by using simple, reliable and non-invasive linear anthropometry measurement techniques are proposed. As a result, the use of expensive surface scanning devices is minimized to model building and simple cost effective linear anthropometry measures can be used for surface anthropometry prediction.

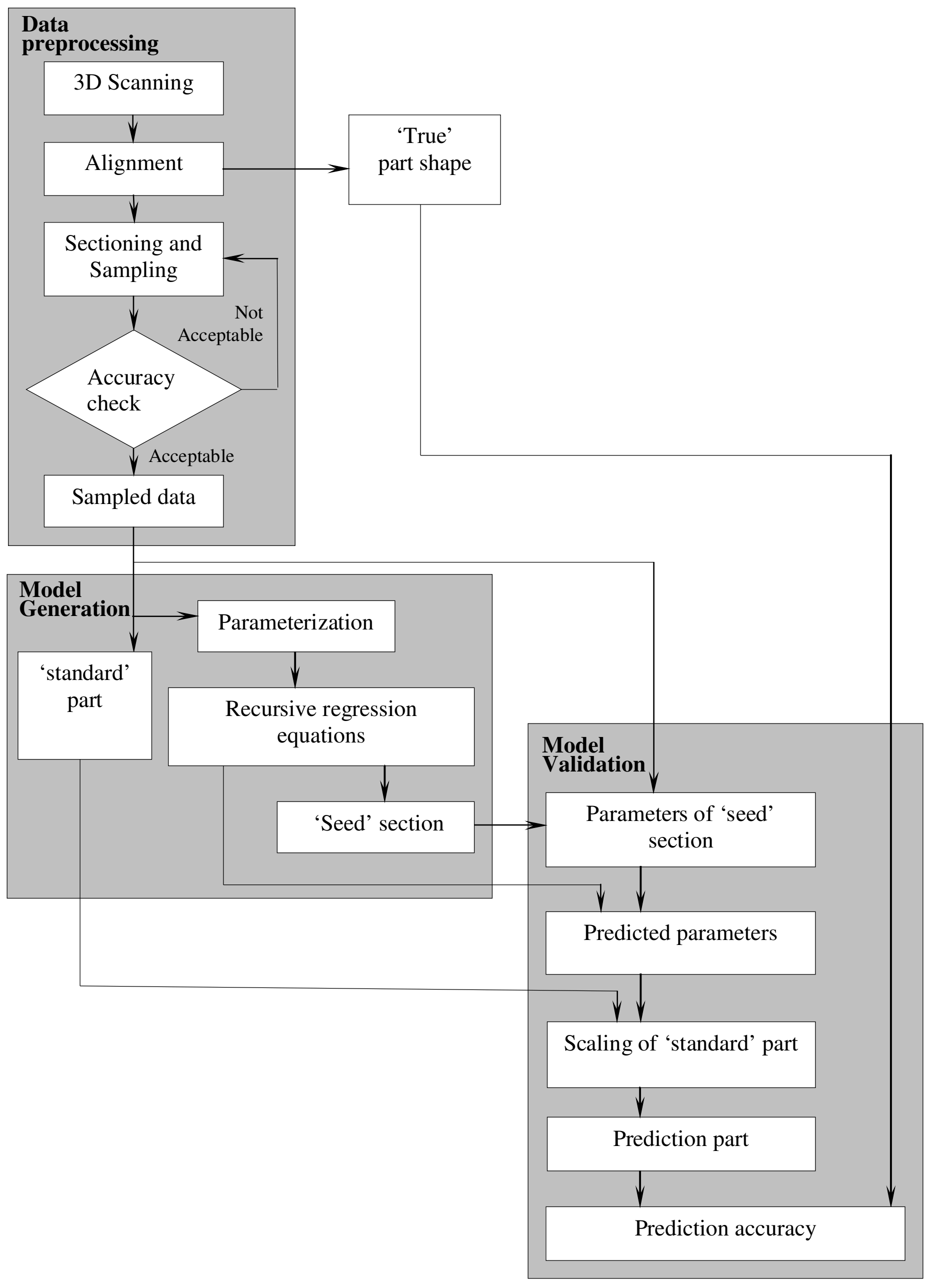


Figure 1. Flowchart of the prediction model.

Surface anthropometry prediction involves spline curve and surface fitting [26], recursive regression equations [27, 28] and scaling techniques [27]. Luximon and Goonetilleke [28] have used the RRET technique and proportional scaling to predict the feet of Hong Kong Chinese to an accuracy of less than 2.5 mm, while using only four anthropometric parameters and a 'standard' foot. Since RRET and scaling techniques seems to provide an accurate method for shape prediction, in this study, a general prediction model with several variations is provided so that more body parts can be predicted using this simple method. The flow chart for the development and validation of the prediction model is shown in Figure 1. The main parts are data pre-processing, model generation and model validation. Data pre-processing includes scanning, and alignment and sampling. Model generation includes creation of a standard shape, section parametization, and generation of regression equations between the parameters. Validation includes checking of accuracy based on a different data set. These are discussed in details in the following sections.

II. Data Pre-processing

A. 3D scanning

For 3D scanning, any type of scanner to capture the external shape of the body (Figure 2) or any specific part can be used. Since a general method to build the prediction model has been proposed, some changes may be required to adopt for specific parts. It is assumed that N_s number of participants is used for the model development. In this formulation, left and right sides of the parts are not distinguished, but during the formation of a specific part, the differences between left and right sides can be included as in Luximon and Goonetilleke [28]. For the i^{th} participant the scanned part has P_i number of points. The points are p_{ik} (where $i = 1,...,N_s$; $k = 1,...,P_i$). The coordinates of the point p_{ik} is (x_{ik}, y_{ik}, z_{ik}).

Figure 2. Laser scanned data

B. Alignment and sampling

Since all the scanned part might not be aligned in the same reference axis, all the parts have to be aligned on a consistent axis. The axis of alignment can be based on some anthropometric landmarks, commonly used axis or based on mathematical and statistical methods (such as principle component methods). For example, for the case of the human foot, heel centre line is commonly used [27-29]. For the arm, leg and body principal component can be used. For head data, eye location can be used for alignment. After alignment, the coordinates of point $_ap_{ik}$ is $(_ax_{ik}, _ay_{ik}, _az_{ik})$ as shown in figure 3. The part is aligned to have the axis with the highest variation along the Z-axis.

Once the part has been aligned, cross sections are extracted perpendicular to the Z-axis, called the 'main' axis. The length of the aligned part along the main axis is L_i (where $i = 1,...,N_s$). Cross-sections perpendicular to the main axis at δ_j (where $j = 1,...,N_{sec}$) of L_i are extracted, where N_{sec} is the total number of cross-sections extracted (Figure 4). δ_j is monotonically increasing with j. The separation between the sampled cross sections need not be uniform, but it has to be consistent between the different participants. The extracted sections are S_{ij} (where $i = 1,...,N_s$; $j = 1,..., N_{sec}$) and the z-value for the sections are given by (1). Then, for each section, a fixed number of points are extracted using different sampling methods [28]. The number of points for section S_{ij} is sN_j. For participant I, the number of points is same. The points after sampling are $_sp_{ijk}$, where $i = 1,...,N_s$; $j = 1,..., N_{sec}$; $j = 1,..., sN_j$.

$$_sz_{ij} = \delta_j * L_i \qquad (1)$$

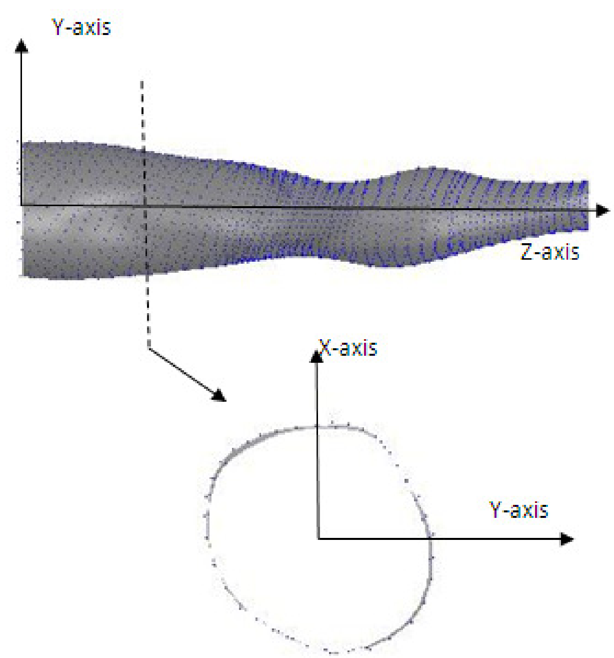


Figure 3. Alignment method

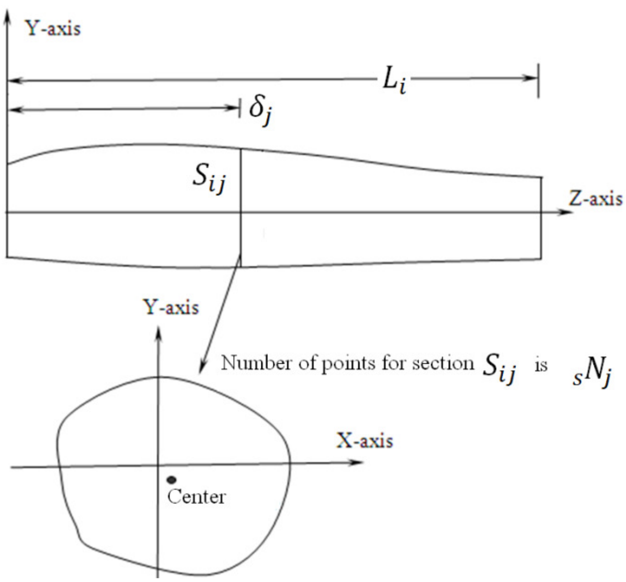


Figure 4. Sectioning and sampling

III. Model Generation

A. 'Standard part'

Some of the part shape data can be used to generate the model, while other shape data can be used for model validation [28]. Assuming that the "standard" part is generated using part shape data of N_m subjects where $N_m < N_s$. The coordinates of the point used to generate the standard part are ($_sx_{ijk}$, $_sy_{ijk}$, $_sy_{ijk}$) where where $i = 1,...,N_m$; $j = 1,..., N_{sec}$; $j = 1,..., sN_j$. The 'Standard' part is the representation of the given part for a given population. There can be several methods to generate the 'standard' part, based on different statistical methods such as geometric mean, arithmetic mean, mode, median, etc. Equations (2), (3), and (4) show the x, y, and z coordinates of the 'standard' part when arithmetic mean is used. The standard foot shape has N_{sec} number of sections.

B. Parametization

$$\bar{x}_{jk} = \frac{1}{N_m}\sum_{i=1}^{N_m} {}_s x_{ijk} \tag{2}$$

$$\bar{y}_{jk} = \frac{1}{N_m}\sum_{i=1}^{N_m} {}_s y_{ijk} \tag{3}$$

$$\bar{z}_{jk} = \frac{1}{N_m}\sum_{i=1}^{N_m} {}_s z_{ijk} \tag{4}$$

The 'standard' part represents the shape of a given population and it can be stored in a database. While the 'standard' part is being developed, parameters can be extracted from the cross sections of the aligned part.

Each cross section can be parameterized using several anthropometric variables [27]. Figure 5 shows some of the parameters that can be used, such as maximum y deviation (H^+), minimum y deviation (H^-), maximum x deviation (W^+), minimum x deviation (W^-), height (H), width (W) and radius (R_θ) at θ degrees and circumference (C). The number of parameterization will determine the accuracy and complexity of the model. Furthermore, anthropometric studies are needed to determine the importance of the different parameters. Goonetilleke et al. [30] and Luximon and Goonetilleke [31] have used principle component and factor analysis to find the relative importance of different foot related parameters.

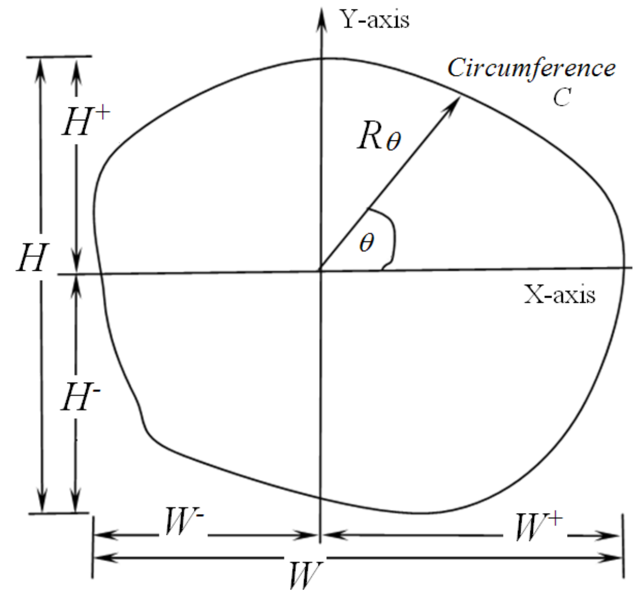


Figure 5. Anthropoemtric parametization

C. Reccursive regression equation

The purpose of the recursive regression equation is to find the relationship of the anthropometric dimensions of all the sections of the part given the anthropometric dimension of one section. For example, one regression equation is build from anthropometric measure height (H) at section i and height at section j. The R^2 values are also recorded. If we have Na anthropometric measures and N_{sec} sections, we can generate N_a x (N_{sec} - 1) equations if we consider consecutive sections. Using these regressions equations, knowledge of one set of values for N_a anthropometric measures ('seed section), we will be able to predict all the N_a x N_{sec} anthropometric measure values. There are a number of ways to find the best 'seed' section and build the regression equations. Luximon and Goonetilleke [28] developed linear regression equations between the anthropometric measures of adjacent sections. The best 'seed' section was found by using different 'seed' section to predict the anthropometric measures and choosing section that provided the highest correlation between the original set of anthropometric measures. For complex models N_a x (N_{sec} -1) x (N_{sec} -2) equations may be needed. This problem can be solved using travelling salesman method [32].

IV. MODEL VALIDATION

The model can be validated using 3D scanned data of a different set of N_v participants where $N_v < N_s$ and $N_v + N_m = N_s$. The model validation involves measurement or extraction of parameters of the 'seed' section, prediction of parameters of all the section based on the 'seed' section, scaling of the 'standard' part. Once the shape is predicted, the prediction error can be calculated when we compare it with the original data.

Once we have the predicted parameters of the sections, the standard part has to be scaled. There can be different scaling methods based on the different parameters. Luximon and Goonetilleke [28] have discussed proportional scaling. If the parameters are orthogonal (such as width and height) then the sections can be scaled independently (figure 6). However, if the parameters are not orthogonal different scaling methods need to be developed. After scaling the predicted shape has coordinates $({}_px_{ij}, {}_py_{ij}, {}_pz_{ij})$.

For participant i the original shape after alignment has coordinates $({}_ax_{ik}, {}_ay_{ik}, {}_az_{ik})$ where $k = 1,...,P_i$. The coordinates of the predicted foot is $({}_px_{ij}, {}_py_{ij}, {}_pz_{ij})$ where $i = 1,...,N_m$; $j = 1,..., N_{sec}$; $j = 1,..., sN_j$. The error is computed based on the shortest distance from the predicted foot to the real foot [33]. Different statistics can easily be calculated to compare prediction accuracy. Error plots are also useful to show the error distribution at different regions [28].

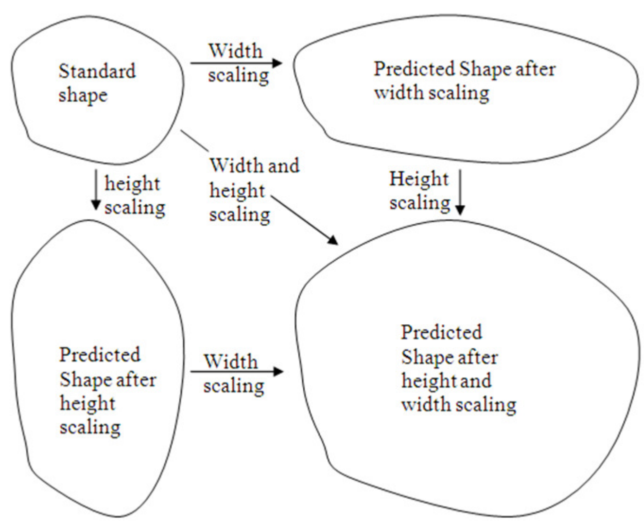


Figure 6. Scaling

V. DISCUSSION

Anthropometric measures are widely used and due to recent advances, there are many types of equipment (3D scanners) to capture surface anthropometry. However, 3D scanners are relatively expensive and are available at few labs. It requires qualified personnel to operate and provides huge number of data points that are sometime difficult to manage. On the other hand, linear anthropometrics have existed for centuries and it is relatively very easy to measure linear anthropometric dimensions. For instance, if we want to buy custom-made shoes through the internet, it is much easier to provide few anthropometric measures (such as length, width, height, heel width, etc) than to scan the foot and provide thousands of points on the surface of the foot. As a result, methods need to be developed to capture the advantages of linear anthropometry as well as surface anthropometry. In this study, a general model was proposed to generate surface anthropometry from linear anthropometry and standard shape. The standard shape can be stored in databases based on age, sex, race and gender. When there is a need to reproduce a 3D shape of a part, there is no need to scan the part. Instead, few anthropometric measures of the part can be used to modify the standard to create accurate 3D shape of a part without the need for expensive scanning. The method has been validated for foot modelling [27, 28] and thus it can be applied to other parts of the body. The model can be modified for specific applications. Results from past studies [28] have shown that recursive regression equations technique (RRET) is a useful technique. The technique has been used for foot shape prediction.

VI. CONCLUSION AND FUTURE WORK

As there are more and more technological innovations, there is a shift from linear anthropometric measures to surface anthropometric data in order to satisfy the ever-changing needs of the society. People are constantly looking for comfortable and 'proper' fitting wearable that not only match the linear anthropometric dimensions but also accommodate the complex surface of the body. In addition, more surface information is needed in medical, archaeological and forensic disciplines. As a result, the linear anthropometric table even though useful is not able to satisfy with the current needs. Thus, in order to have accurate information on body dimensions, surface anthropometry database has to be developed. Since data for linear anthropometric is widely available based on age, race, region, and occupation and methods to capture linear anthropometry are non-invasive, inexpensive, simple, portable and reliable, it is wise to use linear anthropometry to generate surface anthropometry. This study provided a general modelling technique, to convert linear anthropometry to the complex surface anthropometry. Simple recursive regression equations technique and scaling technique were used to build the prediction model. Model building involved data collection, alignment, cross sectioning, point sampling, averaging and regression equations development. Once the model has been built, given few anthropometric measures, the 'standard' part can be scaled to generate a predicted 3D shape. Studies in foot modelling have shown that this method can predict the foot shape accurately (≈ 2 mm) using only 4 parameters [27, 28] including length, width, height and curvature. The accuracy of the predicted shape will generally be better if many anthropometric measures are used. The model parameters can be adjusted to obtain the required accuracy depending on different applications.

Future plan for this study includes modelling different body parts at based on different accuracy level and improvement of this technique by including selected sections. The application of this study is in reconstructive surgery, forensic, anthropology, design, psychology, and other fields involving digital human models.

ACKNOWLEDGMENT

This study was supported by RGC General Research Fund (B-Q26V).

REFERENCES

[1] WHO, Physical Status: The Use and Interpretation of Anthropometry, Technical Report No 854, World Health Organization, Geneva, 1995.

[2] S. Pheasant, Bodyspace: Anthropometry, ergonomics, and the design of work, London: Taylor and Francis, 1996.

[3] A. R. Tilley and Henry Dreyfuss Associates, The measure of man and woman: human factors in design, New York: John Wiley and sons, 2002.

[4] F. E. Johnston, “Anthropometry,” in The Cambridge encyclopaedia of human growth and development, S. J. Ulijaszek, F. E. Johnston and M. A. Preece, Eds. Cambridge: Cambridge University Press, 1998, pp. 26-27.

[5] M. H. Al-Haboubi, “Statistics for a composite distribution in anthropometric studies,” Ergonomics, 40, 1997, pp. 189-198.

[6] W. E. Woodson and D. W. Conover, Human engineering guide for equipment designers (2nd ed.), California: University of California Press Berkeley, 1964.

[7] K. Gielo-Perczak, “The golden section as a harmonizing feature of human dimensions and workplace design,” Theoretical Issues in Ergonomics Science, 2, 2001, pp. 336-351.

[8] A. Ozaslan, M. Y. Iscan, I. Ozaslan, H. Tugcu, and S. Koc, “Estimation of stature from body parts,” Forensic Science International, 132, 2003, pp. 40-45

[9] M. Henneberg, E. Simpson, and C. Stephan, “Human face in biological anthropology: Craniometry, evolution and forensic identification,” in The human face: measurement and meaning, M. Katsikitis, Ed. Boston : Kluwer Academic Publishers, 2003, pp. 29-48.

[10] K. R. Fontaine, G. Gadbury, S. B. Heymsfield, J. Kral, J. B. Albu, and D. Allison, “Quantitative prediction of body diameter in severely obese individuals,” Ergonomics, 45, 2002, pp. 49-60.

[11] O. Giampietro, E. Virgone, L. Carneglia, E. Griesi, D. Calvi, and E. Matteucci, “Anthropometric indices of school children and familiar risk factors,” Preventive Medicine, 35, 2002, pp. 492-498.

[12] N. N. Prasad and D. V. R. Reddy, “Lip-Nose complex anthropometry,” International Journal of Cosmetic Surgery and Aesthetic Dermatology, 4, 2002, pp. 155-159.

[13] R. Z. Stolzenberg-Solomon, P. Pietinen, P. R. Taylor, J. Virtamo, and D. Albanes, “A prospective study of medical conditions, anthropometry, physical activity, and pancreatic cancer in male smokers (Finland),” Cancer Causes and Control, 13, 2002, pp. 417-426.

[14] B. Bogin and R. Keep, “Eight thousand years of economic and political history in Latin America revealed by anthropometry,” Annals of Human Biology, 26, 1999, pp. 333-351.

[15] R. Floud, “The dimensions of inequality: Height and weight variation in Britain, 1700-2000,” Contemporary British History, 16, 2002, pp. 13-26.

[16] T. K. Oommen, “Race, religion, and caste: Anthropological and sociological perspectives,” Comparative Sociology, 1, 2002, pp. 115-126

[17] T. Reilly, J. Bangsbo, and A. Franks, “Anthropometric and physiological predispositions for elite soccer,” Journal of Sports Sciences, 18, pp. 669-683.

[18] P. Tothill and A. D. Stewart, “Estimation of thigh muscle and adipose tissue volume using magnetic resonance imaging and anthropometry,” Journal of Sports Sciences, 20, 2002, pp. 563-576.

[19] M. Westerstahl, M. Barnekow-Bergkvist, G. Hedberg, and E. Jansson, “Secular trends in body dimensions and physical fitness among adolescents in Sweden from 1974 to 1995,” Scandinavian Journal of Medicine and Science in Sports, 13, 2003, pp. 128-137.

[20] C. M. Worthman, “Recumbent anthropometry,” in The Cambridge encyclopaedia of human growth and development, S. J. Ulijaszek, F. E. Johnston and M. A. Preece, Eds. Cambridge: Cambridge University Press, 1998, pp. 29.

[21] S. Ulijaszek, “Measurement error,” in The Cambridge encyclopaedia of human growth and development, S.J. Ulijaszek, F.E. Johnston and M.A. Preece, Eds. Cambridge: Cambridge University Press, 1998, pp. 28.

[22] P. R. M. Jones and M. Rioux, “Three-dimensional surface anthropometry: Applications to the human body,” Optics and Lasers in Engineering, 28, 1997, pp. 89-117.

[23] V. A. Deason, “Anthropometry: The human dimension,” Optics and Lasers in Engineering, 28, 1997, pp. 83-88.

[24] A. D. Linney, J. Campos, and R. Richards, “Non-contact anthropometry using projected laser line distortion: Three dimensional graphic visualization and applications,” Optics and Lasers in Engineering, 28, 1997, pp. 137-155.

[25] D. B. Chaffin Digital human modeling for vehicle and workplace design, Warrendale, PA: Society of Automotive Engineers, 2001.

[26] B. K. Choi, Surface modeling for CAD/CAM: Advances in industrial engineering, Vol. 11, (Amsterdam: Elsevier). 1991.

[27] A. Luximon, Foot shape evaluation for footwear fitting, Unpublished PhD thesis, Hong Kong University of Science and Technology, Hong Kong, 2001.

[28] A. Luximon and R. S. Goonetilleke, 2003, “Foot shape modeling,” Human Factors: The Journal of the Human Factors and Ergonomics Society, 46(2), 2004, pp. 304-315.

[29] R. S. Goonetilleke and A. Luximon, “Foot flare and foot axis,” Human Factors, 41, 1999, pp. 596-607.

[30] R. S. Goonetilleke, C-F. Ho, and R. H. Y. So, “Foot anthropometry in Hong Kong,” Proc of the ASEAN 97 Conference, Kuala Lumpur, Malaysia: ASEAN, 1997, 81-88.

[31] A. Luximon and R. S. Goonetilleke, Dimensions for footwear fitting, Proc of the International Ergonomics Association 2003 conference, Seoul, Korea, 2003, (CD-ROM).

[32] G. Gregory and P. P. Abraham The traveling salesman problem and its variations, Boston : Kluwer Academic Publishers, 2002.

[33] Luximon, R. S. Goonetilleke, and K. L. Tsui, “Foot landmarking for footwear customization,” Ergonomics, 46, 2003, pp. 364-383.

 ISBN: 978-1-61208-250-9

A Modular Interface Design to Indicate a Robot's Social Capabilities

Frank Hegel
Applied Informatics Group (AGAI)
Bielefeld University, Universitätsstraße 21–23
33615 Bielefeld, Germany
e-mail: fhegel@techfak.uni-bielefeld.de

***Abstract* – This paper presents an intersection of human-like appearance, product design, and information design in order to systematically manipulate a robot's conceptual user interface design. The social robot 'Flobi' appears as an iconic cartoon-like character to mediate between users and application scenarios. Flobi's interface design consist of three visual dimensions to choreograph user expectations of the robot's capabilities, traits, and competences. First, the robot has dynamic facial features to display various emotional expressions. Second, the structural head design consist of exchangeable modular parts which are magnetically connected. Through the modular design the visual features of Flobi (e.g., hairstyle, facial features) can be altered easily in order to create various characters. Third, different clothing will prospectively be used to trigger the robot's social roles. All three dimensions of visual features are highly likely to have an effect on the evaluation of Flobi's traits and capabilities.**

***Keywords* – Social Robots, Industrial Design, Human Factors**

I. INTRODUCTION

To date, social robots are mainly designed to engage in social scenarios which are familiar to users. Specifically, robots mainly have to provide a social communicative functionality that the interaction feels natural and intuitive to those who interact with the robot. In this respect, a social robot is generally a specific kind of interface metaphor in order to provide human-like interaction patterns [1].

Roboticists usually design robots to appear lifelike in order to represent natural (i.e., familiar) behaviors. On the one hand, some social robots strongly appear human-like, because the goal is to create android surrogates of existing persons. On the other hand, many social robots appear rather zoomorphic, caricatured, or technical [2]. But even if social robots vary greatly in terms of appearance their aesthetic form is primary intended to follow a specific familiar function, because this likely enhances the machines' comprehensibility to guarantee an intuitive usage. To illustrate, a social robot acting in a kitchen might be designed like a kitchen aid whereby a robotic learning aid might rather be shaped to appear like a teacher in order to mediate a certain degree of expertise.

One possibility of displaying different functionalities by a robot's appearance is to consider exchangeable visual features which represent aspects of a specific expertise. Regarding the industrial design of the robot Flobi, exchangeable visual features have been conceptualized and implemented [3]. As depicted in Figure 1, Flobi generally appears human-like due to its facial features. These features were implemented because previous research has demonstrated that a certain degree of human-likeness in appearance is necessary to produce facial expressions that are intuitively understandable. In addition, human-like appearance increases likely a robot's predictability in terms of human-like behaviors.

Taken together, the design of Flobi consists of three dimensions: First, the implementation of unambiguously recognizable facial expressions. Second, a modular surface-design makes it possible to quickly alter Flobi's visual character. Third, dress codes are considered to signify the robot's social roles in order to increase the predictability of the its expertise.

Figure 1: The social robot Flobi

After outlining related research in Section II, I present the key aspects of Flobi's visual conceptualization in Section III. Section IV discuss the visual conceptualization and concludes this article.

II. RELATED WORK

In this chapter, I introduce into research on the appearance of social robots and related effects of such visual representations. Moreover, I give a short introduction into relevant principles of iconic communication with regard to information design. These principles illustrate how specific information can visually be implement into objects as well as interfaces.

 ISBN: 978-1-61208-250-9

2.1 Appearance of Social Robots

Originally, the meaning of the term ‘automaton’ implies autonomous beings having the ability to move on their own. For instance, Vausanson’s flute and tabor player and Wolfgang von Kemepelen’s famous chess player, ‘The Turk’, designed in the mid 1700s, are early encounters between lifelike forms and machines. Both machines invoked on people’s projections and expectations due to the behavior displayed by their lifelike form. Even today, social roboticists connect form and function in an attempt to develop lifelike social robots [4]. Thus, one general objective of social robotics research is to create robots that engage in social scenarios which are familiar to users. Given this objective, robots have to provide a social communicative functionality that is natural and intuitive to those who interact with the robot.

Social robots vary greatly in terms of appearance to indicate specific behaviors or applications. Some social robots appear highly anthropomorphic while others appear rather zoomorphic, caricatured, or functional. According to [2] an anthropomorphic appearance is recommended to support meaningful interactions [5] with users because many aspects of nonverbal communication are only understandable if expressed in similarity to a human-like body. For instance, emotional displays are highly iconic to emotional displays of human beings. Zoomorphic robots are intended to look like their animal counterparts to support the idea that an observer expects the robot to behave like an animal. In some cases this might be helpful to communicate the functional limitations of a social robot. To illustrate, a dog only partially understands human speech, but maybe this represents today’s recognition rates of current speech recognition software [6]. Robots with a caricatured appearance are mainly designed both to not elicit any expectations based on familiarity and to focus on specific attributes like mouth (i.e., speeking) or eyes (i.e., seeing). Finally, functional shaped robots are designed in a technical manner to illustrate their ultimate technical functions. This functional approach corresponds in a certain respect to the claim by Sullivan [7] that ‘form ever follows function’. In this case, the robot-designer expects that the user is able to understand capabilities of the robot by looking at its technical features.

Nevertheless, an human-like appearance matches probably best with the idea to implement artificial human-like behaviors to support intuitive interaction patterns. But in how far does an anthropomorphic appearance of robots affect the assessment of them? Specifically, recent research has shown that human-likeness of agents relates highly to the phenomenon of anthropomorphism.

2.1.1 Anthropomorphism

Anthropomorphism entails attributing human-like characteristics, properties, or mental states to real as well as imagined non-human agents and objects [8]. According to the familiarity hypothesis [9], people draw anthropomorphic inferences, because it allows us to explain things we do not understand in terms that we do understand – and we best understand ourselves as human beings.

However, the ‘Three-Factor Theory of Anthropomorphism’ [8] claims that the extent to which people anthropomorphize is determined by three general factors: First, *Effectance Motivation* describes the need to interact effectively with one’s environment. Attributing human characteristics and motivations to non-human agents increases the ability to make sense of an agent’s actions and consequently reduces uncertainty. Second, *Sociality Motivation* describes the need and desire to establish social connections. To illsutrate, when people feel lack of social connection they anthropomorphize objects more strongly. They do so to satisfy their need for affiliation. Finally, *Elicited Agent Knowledge* serves as a basis for induction primarily because such knowledge is acquired earlier and is more detailed than knowledge about non-human agents or objects. As a result, the more human-like an object appears, the more do people probably use themselves as a source of induction when judging non-human agents.

The key role of human-likeness in appearance has been demonstrated in an experiment by [10]. These authors conducted an fMRI study with three different robot targets which differed in their degree of human-like appearance. Participants’ brain activity was measured during playing an interactive game (.i.e., Prisoners' Dilemma) with these robots. The results showed that the degree of human-likeness had significant effects on the participants’ cortical activities associated with Theory of Mind (ToM) and their judgments of the different robots. Summing up, the more human-like an interaction partner appears, the more do participants speculate implicitly about the robot’s intentions. Moreover, it has been shown that the quantity of facial features implemented in a robot affect perceptions of human-likeness [11]. But in addition, faces imply various qualities which have effects on judgements. Specifically, due to the fact that a face continuously conveys information especially poor designed faces may cause negative attitudes towards an artificial agent or a robot [12].

2.1.2 Qualities of Facial Appearance

A vast body of research has shown that people attribute more positive traits to attractive people than to unattractive ones. [13-16]. To illustrate, attractive humans are commonly judged as warmer, kinder, stronger, more sensitive, interesting, poised, modest, sociable, and outgoing [13]. Moreoever, even babies prefer playing with attractive puppets [17]. Importantly, [18] suggested that an attractiveness bias is also applicable to objects. Further, there is evidence that unattractive objects can elicit uneaseness. This phenomenon is wellknown as the ‘uncanny valley’ hypothesis [19]. Such attributions are particularly true for faces with abnormal facial features (e.g., with regard to a bigger eye size of 150%). This suggests that the human visual system is particularly sensitive to cues indicating human-likeness [20].

However, not only does attractiveness or human-likeness of a face influence social perceptions. This is also true for babyfacedness. People with babyfaced facial features (e.g., curved forehead, large eyes, small nose and chin) are characterized as warmer, more naive, submissive, less dominant, and less competent than mature-faced counterparts [21]. Regarding Flobi’s appearance, we designed the head babyfaced in order to both facilitate human-robot interactions and display the limitations regarding the robot’s limited skills. Additionally, a modular design has been realized in order to modify the users’ evaluations of the robot's capabilities. The concept of how to systematically implement modular parts was mainly inspired by principles of information design and iconic communication.

2.2. Iconic Communication

To date, iconic communication systems often become valuable as new information systems call for designs that cut across language barriers. Therefore, iconic communication is an interesting source of inspiration when designing novel electronic interfaces such as social robots. Principles of iconic communication are universal and they are almost not tied to unique features of a particular language or culture [22].

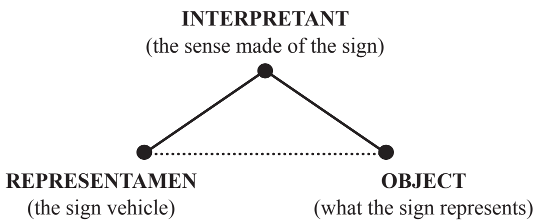


Figure 2: Sign relations according to Peirce. The dotted line means that there does not have to be any direct relation between the form of the sign and what it stands for

Generally, any information is mediated by signs and the concept of signs is essential to understand iconic communication: "A sign... [in the form of a representamen] is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the interpretant of the first sign. The sign stands for something, its object. It stands for that object, not in all respects, but in reference to a sort of idea, which I have sometimes called the ground of the representamen" [23] (see Figure 2). Furthermore, Peirce distinguished three general types of representamen: First, *indexical* signs denote their objects by virtue of an actual (physical) connection involving them (e.g., footsteps are connected to a living creature, smoke is connected to fire). Second, *symbolic* signs are connected to the represented objects by convention (e.g., words or language in general). Third, *iconic* signs represent objects by their similarity (e.g., photograph, natural drawing). Furthermore, Peirce then again divided icons into three subtypes: images, diagrams, and metaphors. Evidently, icons are naturally related to visual communication.

By using the term 'iconic communication system' I refer to a systematical combination of several icons to derive more specific meanings. To illustrate, I introduce into one specific aspect of the 'International System Of TYpographic Picture Education (ISOTYPE).

ISOTYPE [24] was basically invented in the 1920s by the sociologist Otto Neurath and the designer Gerd Arntz to visualize social and economical facts particularly with regard to facilitate the understanding of complex data for less educated people. Beyond that, ISOTYPE was intended to support foreign people having intuitively access to specific knowledge. To illustrate, icons at airports are richly implemented to help foreign

Figure 3: In ISOTYPE ethnic groups are demonstrated systematically by various iconic head types

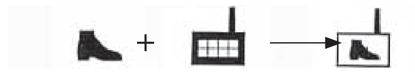


Figure 4: In ISOTYPE a combination of the visual items 'shoe' and 'factory' means a 'shoe factory'

people finding luggage, toilets, exits etc. With the invention of ISOTYPE there has been established new standards for presenting data and it fundamentally influenced several topics in the field of information design and interface design as well.

Neurath and Arntz created a limited number of icons for international use that should be readable without using further descriptions. At a first glance, an icon created ISOTYPE displays only the most important details of the represented object – in a way that it is just identifiable. At a second glance, visual attributes are used to elicit specific meanings of the object. To demonstrate, in Figure 3 there are circles representing human heads and, most importantly, different hats representing ethnic groups.

Such combined icons can be used to signify their relations in order to initiate more specific meanings. To illustrate, in Figure 4 a visual combination of 'shoe' and 'factory' simply is a 'shoe factory'. Such combined icons are frequently used in the field of interface design. For instance, *Susan Kare,* who created the interface of the *Macintosh Operating System* [25], differentiated on a first dimension between a general icon representing 'documents' and a prictograph representing 'applications' (see Figure 5). On another second dimension she connected additional 'task-icons' (e.g., text, draw, paint) to the documents and applications to categorize specific types of them. This schema supports users to quickly recognize the interface's objects. According to these concepts of combined icons, three dimensions of exchangeable visual parts were implemented to implicitly choreograph the users' perceived capabilities of the robot.

III. VISUALIZATION OF FLOBI

According to previous research [11], the extent to which a face is perceived human-like depends on the quantity of facial features. To realize a high degree of human-likeness, Flobi's facial features include eyes, eyelids, lips, ears, eyebrows. However, with regard to qualities of these features, Flobi appears cartoon-like with a certain similarity to humans to trigger natural interactions. Additionally, the robot appears babyfaced that people potentially

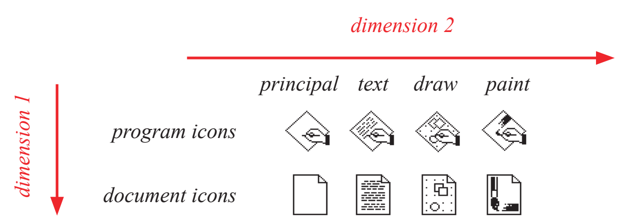


Figure 5: Icons of the first Apple Macintosh System

 ISBN: 978-1-61208-250-9

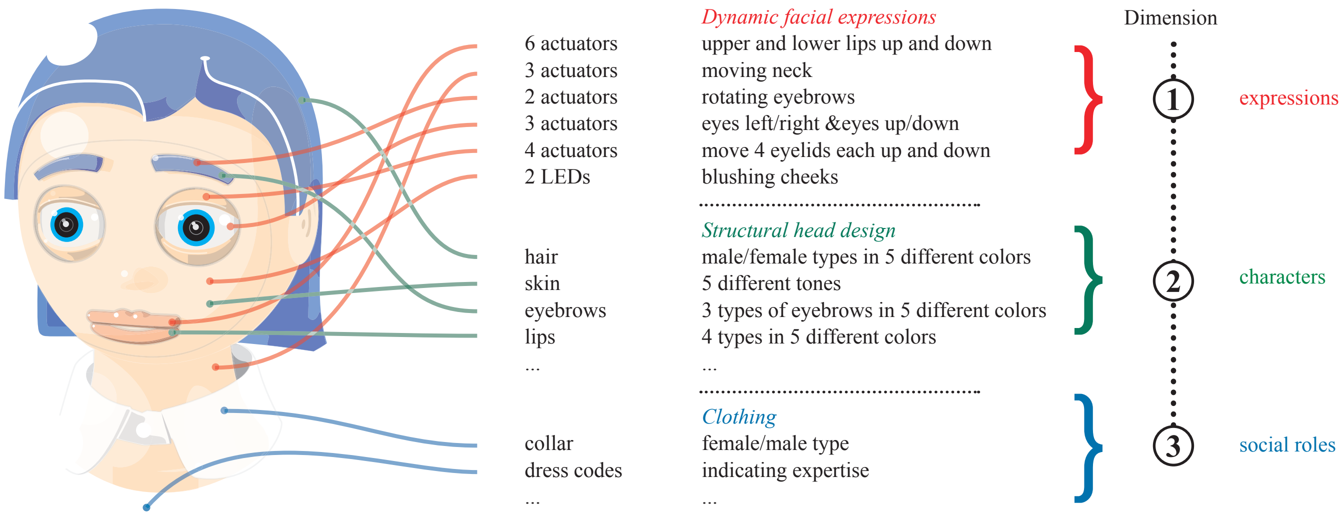


Figure 6: Iconic Communication with three dimensions of dynamic features

judge the robot to be submissive, more naive, and social warmer. Consequently, Flobi has relatively large eyes, a small chin, small lips, a small nose, and a round head shape similar to a baby's head [21]. Moreover, no technical junctions are visible in order to have a consistent character design and to avoid the aforementioned 'uncanny valley' [20].

Additionally, the size of the facial features is meant to signify a match between form and the product's actual function: The user shall infer Flobi's capabilities from its appearance. To give a visual indication of the robot's functional capabilities, Flobi has emphasized large eyes due to comparatively good visual capabilities, normal ears due to available hearing capabilities, and a small nose to indicate that Flobi does not have any olfactory capabilities. Furthermore, characters in computer games often have large eyes to improve the readability of attention – the direction of large eyes compared to small eyes is more likely to be recognized [26].

One of the significant issues that arises when we want to match the users' expectations with regard to different application scenarios is the conceptualization of exchangeable visual cues which can be altered dynamically. Specifically, Flobi has principally been developed to engage within various scenarios. Currently the robot is used in an emotional scenario [27] and in a sports scenario to support astronauts performing their daily physical training [28]. In order to use the robot in further scenarios, an iconic system of three dimensions was conceptualized (see Figure 6): First, the robot has the capability to dynamically display various facial expressions. Second, all perceivable parts of Flobi's head design are easily exchangeable. Finally, visual cues of dress codes will be used to initiate schemata of expertise.

3.1 Dimension 1: Displaying Facial Expressions

Altogether, Flobi has 18 degrees of freedom to display facial expressions, such as basic emotions like happiness, sadness, fear, surprise, and anger. Four actuators move the upper and lower eyelids, two actuators rotate the eyebrows, three actuators move the eyes, three actuators move the neck, and finally, six actuators were implemented to animate the robot's lips. Furthermore, by means of four LEDs, red or white light can be projected onto Flobi's cheek surfaces in order to prospectively indicate either shame or healthiness. Displaying shame is an interesting feature, because it signifies a uniquely human emotion and has not yet been investigated deeply in robotics.

Because product design requires covering technical conjunctions, we developed a 'hole-free' robot head. It was particularly challenging to meet this requirement with regard to Flobi's lip movements. In few robots LED technology is used to project lips onto the face. Nevertheless, we decided against LED technology to display the robot's lips because of the unnatural appeal of LEDs. Instead, Flobi's upper and lower lips consist of neodymium magnets that can be actuated separately. Behind the robot's mask, coupled magnets are actuated on sliding axes, with the motion range overlapping between upper and lower lips. The large and overlapping motion range makes it possible to realize a relatively natural facial expression, because the corners of Flobi's mouth are not fixed. Flobi differs from some prominent social robots in this regard (e.g., the Philips iCat [29]), because in these robots, the corners of the mouth are commonly fixed. Flobi's lip actuators can lift the corners of the mouth to form natural smile without exposing holes or hardware.

In a first study regarding the readability of emotional displays, it has been shown that participants are able to distinguish a set of signified facial expressions. The face with its actuated features makes it possible to display basic emotions as happiness, sadness, fear, surprise and anger (see Figure 7). However, are users able to classify these distinct emotional states correctly? This was tested in an online survey [31] with 259 participants (160 female, 90 male) who evaluated the emotional

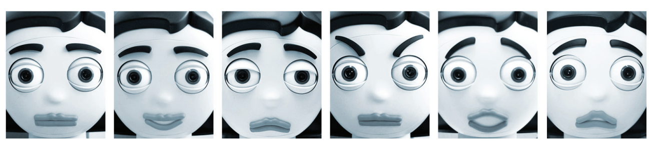


Figure 7: Emotional displays of Flobi. From left to right: neutral, happiness, sadness, anger, surprise, and fear

displays of Flobi. The participants were ranging in age from 17 to 67 years, with a mean age of 27.7. Conceptually replicating [30], participants were presented with five images of the robot in a randomized order. The images depicted five basic emotions displayed by the robot. Subsequently, participants had to indicate which of the emotions would be portrayed by the robots.

The results demonstrated that Flobi's displays were almost readable: The participants were able to classify displays of sadness (99,2%), happiness (83,3%), anger (81,2%) and surprise (54,5%) relatively correct. The display of fear (33,5%), however, was recognized less well, because obviously, many participants mistook surprise for fear (51,2%). Moreover, it is likely that these displays will be even more readable when they are presented in a certain context. Taken together, by means of this 18 mechanically actuated features Flobi is capable of displaying a variety of meaningful expressions that represent the product's communicative states and intentions.

3.2 Dimension 2: Modular Head Features

Psychological research has shown that even subtle visual cues can lead to a target's categorization, for example, in terms of the person's age, race, or gender. This categorization occurs automatically within individuals and activates knowledge structures such as stereotypes or social roles. Specifically, stereotypes and social roles imply certain sets of behaviors that are expected of a person [32]. For this reason, Flobi's consist of exchangeable modular parts (see Figure 8). Because the modular approach makes it possible to alter Flobi's appearance quickly and flexibly, this robot can theoretically be used as research platform to study HRI in a wide range of different contexts.

By means of the modular conceptualization of Flobi's head, it is possible select specific features to manipulate users' expectations and perceptions of the robot. Practically, all modules of the robot head can be combined as needed. This is possible because most of the head's features are attached to the core by means of neodymium magnets. To build one specific character a set of ten parts in total is required.

Two main parts, a front head and a front neck, are screwed onto the technical core (see Figure 8-6). The face and the back part of the head are connected to the front head part using neodymium magnets. Flobi's back neck is connected to its front neck part using magnets as well. All these parts are available in human-like skin tones of varying shades (see Figure 8-2). The hair parts as well as the upper and lower lips are connected to the head using magnets, too. Flobi's eyebrows can be plugged into an actuator behind the face mask.

In a preliminary study [33] regarding the variable modules, we found that different hair types affect stereotypically knowledge structures in subjects. 60 participants were tested, ranging in age from 19 to 38 years. We used two different hair modules to create a long-haired 'feminine robot' and a short-haired 'masculine robot' (see Figure 9). As predicted, the long-haired 'female' version of Flobi was perceived as more feminine than the short-haired one. The participants were then asked to evaluate the 'gendered' robots in terms of gender-stereotypical traits and the robot's suitability for typically female vs. male tasks. In this manner, the participants rated the robot on a 7-point Likert scale with regard

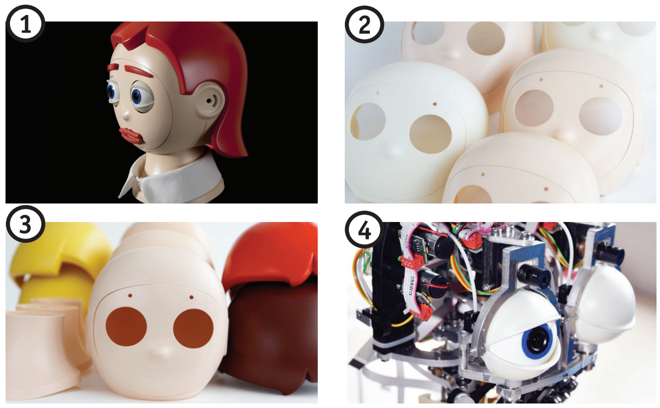


Figure 8: Exchangeanble parts of the robot's head: (1) female Flobi combined with red hair, (2) skin tones, (3) a set of different head parts, (4) technical corpus

to six stereotypically female traits (e.g., friendly, trusting, polite) and six stereotypically male traits (e.g., authoritative, aggressive, dominant). The results showed that the male robot was perceived as possessing more stereotypically masculine traits than the female robot whereas the female robot was perceived as slightly warmer than the male robot.

Furthermore, the participants evaluated the robots' suitability for pretested typically female tasks (childcare, household maintenance, after-school tutoring, patient care, preparing meals and elderly care) and typically male tasks (transporting goods, repairing technical equiptment, guarding a house, steering machines, handcrafting and servicing equiptment). The results show that typically female tasks were perceived as more suitable for the feminine robot relative to the masculine target – and vice versa. Participants perceived the female robot as being more suitable for stereotypically 'female tasks' than the male robot. Vice versa, the male robot was perceived as being more suitable for 'male tasks' [more detailed information regarding this experiment in 33]. Taken together, due to the modular design (in this case a hair module), we are able to create distinct robot characters whose gendered appearance affect the participants' expectations about the robots' personality and expertise.

3.3 Dimension 3: Dress Codes

According to Goffman [34] life in society is a sort of 'theater'. That is, there is a connection between the kinds of acts that people put on in their daily life and theatrical performances. In daily interactions as well as in theatrical performances the 'actors' (individuals) are on 'stage' in front of their audiences. However, people display their social roles not only behavioral, in addition they automatically use cues such as dress codes by which other people (audiences) are able to identify their specific

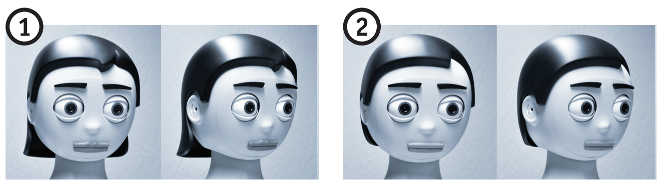


Figure 9: Flobi with female (1) and male (2) hair module

 ISBN: 978-1-61208-250-9

social roles. Commonly, people wear a certain set of clothes that signify their social role. This has been evolved due to different requirements of specific domains. To illustrate, most sports and physical activities are practiced wearing specific clothing, for practical, for comfort, or for safety reasons. In addition, clothing performs a range of social and cultural functions, such as individual, occupational and sexual differentiation, and social status. Probably, in almost all societies, dress codes reflect standards of modesty, religion, gender, and social status. Dress codes may additionally function as a sort of adornment and an expression of personal taste or style. However, triggering role-specific information is likely possible by considering role-specific dress codes.

Role specific information such as clothing codes have already been applied to virtual agents in order to alter a user’s expectations. Virtual agents wearing role-specific clothing were expected to behave appropriate in terms of that roles [35]. Therefore, it is highly likely that context-specific dresses in social robotics also have an effect regarding the assessment of robots as well. To investigate effects of dress codes in robotics, we drew a set of sketches (examples in Figure 10) which will be tested experimentally in order to have knowledge how to choreograph the perceived expertise of a robot by using such codes.

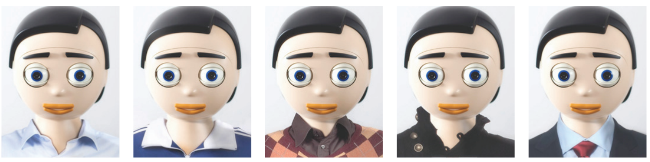


Figure 10: Examples of dress codes applied to Flobi

IV. DISCUSSION & CONCLUSION

Flobi is a cartoon-like social robot with a certain degree of human-likeness. In order to activate natural interaction patterns, one essential feature of Flobi is that it is capable of displaying distinctive meaningful facial expressions. Participants in a pilot study were intuitively able to almost distinguish among five displays of basic emotions. Furthermore, according to principles of information design Flobi’s head consists of exchangeable modular parts to instantly arrange different characters. This set of characters was conceptualized to initiate expectations in terms of stereotypes. Regarding the modularity, it has been shown by means of statistical analysis that even minimal hair cues led people to judge the robot differently. An expansion of another head parts including additional visual attributes is work in progress. In addition to exchangeable modules, different clothings are conceptualized to indicate Flobi’s social role with regard to the its expertise.

Taken together, there currently exist three visual dimensions of dynamic features to modify Flobi’s appearance: (a) movable facial expressions, (b) modular head parts, and (c) exchangeable clothings.

Fundamentally, this concept of exchangeable visuals is not a novel one since it has successfully been implemented in many products such as cell phones (e.g., different skins of preference), cars, and toys (e.g., Playmobil® figures consist of different clothings as well as facial features to indicate specific scenarios of play). In contrast to today´s commercial products, this concept has never been applied to robotic products with the ability of social interaction before – even though it is an interesting feature due to the fact that there is probably an open field of various unknown application scenarios.

The option to exchange visual features enables us to systematically do research on the appearance of robots. Most of today’s social robots have a fixed appearance that allows researchers only to investigate explicitly one specific character. By contrast, the modular design makes it possible to investigate dozens of differently created characters in order to understand the effects of appearance in this field of robotics.

Moreover, a modular robotic design involves potential users into the process of design and enables them to arange their own characters of aesthetical preference. It is likely that people have a bias towards specific characters. To illustrate, some people generally prefer to have female agents while others might prefer male ones. This way, the method of modularity is a first step towards giving the people the ability to build their own enjoyable robotic characters.

The conceptualization of the modular design was mainly realized to indicate familiar knowledge structures in order to choreograph people’s mental models with regard to the robot’s apparently capabilities. To illustrate, a social robot whose job is to support technical maintenance will probably be perceived as having more expertise if the robot has typically a male appearance wearing a repairman’s clothing. This suggestion is supported by our first findings that even the exchange of minimal hair cues affect the user’s perception. However, regarding the modular concept there are three open questions: First, the results of our experiment are currently limited to first impressions of the robot. Accordingly, it might be possible that individuals change their attitudes toward robots due to iterative interactions. Additionally, we are not able to predict how the robot’s actual behavior affects the expectations people have due to their first impressions. Second, with regard to dress codes we first have to conduct experiments whether people ascribe or not specific capabilities due to such codes. Third, the presented studies in this article are limited to the evaluation of images displaying robots. The embodiment of a product might have an effect on the perception and the judgement of social robots as well.

Unfortunately, from an engineering point of view, Flobi is currently far away from being a commercial product. Up to now, two hardware prototypes have been realized mainly to test the functional issues in terms of technical capabilities – a third prototype is work in progress. Therefore, it has not been tested to which extent the aesthetical modularity of robots might be effective in the field of commercial entertainment robots as well. Therefore, a vast body of additional research is needed to draw further conclusions with regard to the engineering, aesthetical, and practical aspects of the modularity concept.

V. ACKNOWLEDGEMENTS

I gratefully acknowledge partial support by the German Aerospace Center (DLR) with funds from the Federal Ministry of Economics and Technology (BMBF) due to resolution 50RA1023 of the German Bundestag.

VI. REFERENCES

[1] F. Hegel, C. Muhl, B. Wrede, M. Hielscher-Fastabend, and G. Sagerer (2009). 'Understanding Social Robots', in the Proceedings of the Second International Conferences on Advances in Computer-Human Interactions (ACHI), Cancun, Mexico.
[2] T. W. Fong, I. Nourbakhsh, and K. Dautenhahn (2002). 'A Survey of Socially Interactive Robots: Concepts, Design, and Applications', Robotics and Autonomous Systems, 42 (3–4), 142–166.
[3] F. Hegel, F. Eyssel, and B. Wrede (2010). 'The Social Robot Flobi: Key Concepts of Industrial Design'. In 19th IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN). Viareggio, Italy
[4] Duffy BR, Joue G (2004). 'I, robot being'. In proceedings of the Intelligent Autonomous Systems Conference (IAS8) 10-13 March, Amsterdam, NL.
[5] C. Breazeal (2002). 'Designing Sociable Robots'. Cambridge, MA: MIT Press.
[6] C. Bartneck (2003), 'Interacting with and embodied emotional Character'. Proceedings of the DPPI2003 Conference, Pittsburgh, 55-60.
[7] L. Sullivan (1896). 'The Tall Office Building Artistically Considered'. Published in Lippincott's Magazine.
[8] N. Epley, A. Waytz, and J. T. Cacioppo (2007). 'On Seeing Human: A Three-Factor Theory of Anthropomorphism'. Psychological Review 114, 864–886.
[9] S. E. Guthrie (1997). 'Anthropomorphism: A Definition and a Theory'. In: R. W. Mitchell et al., 'Anthropomorphism, anecdotes, and animals'. Albany: State University of New York Press., USA.
[10] S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher (2008). 'Can machines think? Interaction and perspective taking with robots investigated via fMRI'. In PLoS ONE, 3(7), 2008. doi:10.1371/journal.pone.0002597.
[11] C.F. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler (2002). 'All robots are not created equal: the design and perception of humanoid robot heads', in DIS '02: Proceedings of the Conference on Designing interactive systems, 321-326, New York, NY, USA.
[12] J.S. Donath (2001). 'Mediated Faces', in the Proceedings of the 4th International Conference on Cognitive Technology, 373-390.
[13] K. Dion, E. Berscheid, and E. Walster (1972). 'What is beautiful is good'. Journal of Personality and Social Psychology, Vol. 24 (3), 285-290.
[14] N. Ambady and R. Rosenthal (1993). 'Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness'. Journal of Personality and Social Psychology, 64, 431-441.
[15] K. Dion (1973). 'Young children's stereotyping of facial attractiveness'. Developmental Psychology, 9, 183-188.
[16] R.L. Dipboye, R.D. Arvey, and D.E. Terpstra (1977). 'Sex and physical attractiveness of raters and applicants as determinants of resume´ evaluations'. Journal of Applied Psychology, 4, 288-294.
[17] J.H. Langlois, L.A. Roggman, and L.A. Ricser-Danner, (1990). 'Infants' Differential Social Responses to Attractive and Unattractive Faces'. Developmental Psychology, 26, 153-159.
[18] D.A. Norman (2004). 'Emotional Design: Why We Love (Or Hate) Everyday Things'. New York: Basic Books.
[19] M. Mori (1997). 'The Buddha in the Robot'. Charles E. Tuttle Co.
[20] J. Seyama and R.S. Nagayama (2007). 'The uncanny valley: Effect of realism on the impression of artificial human faces'. Presence. Teleoperators and Virtual Environments, 16(4), 337-351.
[21] L.A. Zebrowitz (1997). 'Reading faces: Window to the soul?', Boulder, CO: Westview Press, USA.
[22] E. R. Tufte (2001). 'Envisioning Information'. Cheshire, CT, Graphics Press.
[23] Peirce, Charles Sander: Collected Papers of Charles Sanders Peirce (Volumes I-VI ed. by Charles Hartshorne and Paul Weiss, 1931-1935). Cambridge, Massachusetts: Harvard University Press.
[24] O. Neurath (2010). 'From Hieroglyphics to Isotype: A Visual Autobiography'. Princeton Architectural Press.
[25] http://www.kare.com (06/2011).
[26] K. Isbister (2006). 'Better game characters by design: A psychological approach'. The Morgan Kaufmann Publishers Inc., San Francisco, USA.
[27] http://www.sfb673.org/projects/C2/ (12/2012).
[28] http://ekvv.uni-bielefeld.de/blog/uninews/entry/robots_as_fitness_coaches_for (12/2012).
[29] http://www.hitech-projects.com/icat/ (12/2012)
[30] P. Ekman, W.V. Friesen, M. O'Sullivan, I. Diacoyanni-Tarlatzis, R. Krause, T. Piticairn, K. Scherer, A. Chan, K. Heider, W.A. LeCompte P.E. Ricci-Bitti, M. Tomita, and A. Tzavaras (1987). 'Universal and cultural differences in the judgement of facial expressions of emotion'. In the Journal of Personality and Social Psychology 53 (4), 712-717.
[31] F. Hegel (2010). 'Gestalterische Konstruktion eines sozialen Roboters', Dissertation, Faculty of Technology, Bielefeld University, Germany.
[32] S.T. Fiske and S.E. Taylor (1991). 'Social cognition'. Second Edition, McGraw-Hill, New York, USA.
[33] F. Eyssel and F. Hegel (2012). '(S)he's got the look: Gender stereotyping of social robots'. Journal of Applied Social Psychology 42 (9).
[34] E. Goffman (1959). 'The presentation of self in everyday life'. Anchor Books, New York, USA.
[35] B. Hayes-Roth, H. Maldonado, and M. Moraes (2002). 'Designing for diversity: Multi-cultural characters for a multicultural world'. In Proceedings of IMAGINA 2000, Monte Carlo, Monaco, 207-225.

Automatic Discrimination of Voluntary and Spontaneous Eyeblinks.

The use of the blink as a switch interface

Shogo Matsuno, Minoru Ohyama, Shoichi Ohi
Graduate School of Information Environment
Tokyo Denki University
Tokyo, Japan
e-mail:12jkm25@ms.dendai.ac.jp

Kiyohiko Abe, Hironobu Sato
Department of Network and Multi-Media Engineering
Kanto Gakuin University
Kanagawa, Japan
e-mail:abe@kanto-gakuin.ac.jp

***Abstract*— This paper proposes a method to analyze the automatic detection and discrimination of eyeblinks for use with a human-computer interface. When eyeblinks are detected, the eyeblink waveform is also acquired from a change in the eye aperture area of the subject by giving a sound signal. We compared voluntary and spontaneous blink parameters obtained by experiments, and we found that the trends of the subjects for important feature parameters could be sorted into three types. As a result, the realization of automatic discrimination of voluntary and spontaneous eye blinking can be expected.**

Keywords— computer interface; automatic discrimination; voluntary eye blink; spontaneous eye blink

I. INTRODUCTION

The relation between eyeblinks and cognitive states has been pointed out by psychological experiments [1][2]. Realization of automatic discrimination of spontaneous and voluntary eyeblinks is desired. However, it is difficult to automate the detection of eyeblinks or the discrimination of kinds of eyeblinks, because the generation of eyeblinks has large variations between individuals. Therefore, until now, the detection or discrimination of eyeblinks has been carried out manually in most cases. Now, researchers are advancing their studies by aiming at automation of the detection and discrimination of eyeblinks while bearing a computer interface in mind.

A technique using the electro-oculogram (EOG) has been proposed as an automatic detection method for voluntary eyeblinks [3][4]. The EOG method consists of sticking an electrode on the skin near the eyeball and detecting changes in cornea-retina potential. However, adopting the EOG method as a common interface is accompanied with the difficulty of directly equipping the skin with an electrode and exclusively using a special machine.

In this research, the videotape recording (VTR) method is adopted, whereby eyeblinks are measured from a video image. Until recently, automatic blink detection using the VTR method was difficult, because the accuracy was lowered by a shortage of sampling points. Then, a technique for using an interlaced picture was incorporated, which divided the field that others had proposed. Hence, even if we use an ordinary NTSC video camera, we can obtain twice as much time resolution as with natural NTSC video. In this research, we tried to detect a blink waveform by analyzing interlaced pictures taken with an ordinary NTSC video camera and then conducting a discrimination experiment on spontaneous eyeblinks and voluntary eyeblinks. In this experiment, we performed an estimation experiment on automatic detection of eyeblinks and extraction experiments on the shape feature parameters of eyeblink waveforms (two-pattern condition). The results are reported herein.

II. AUTOMATIC DETECTION OF EYEBLINK WAVEFORM BY VTR METHOD

If the time evolution of the eyeblink process is correctly measurable, it is possible to express an eyeblink as a waveform (a so-called eyeblink waveform). In order to identify the kind of eyeblink that has occurred, sampling the eyeblink waveform is first necessary. The typical techniques for sampling eyeblink waveforms are the EOG method and the VTR method. The former technique involves sticking an electrode on the skin near the eyeball, and it acquires the eyeblink waveform by recording the changes in cornea-retina potential. Until now, this has been the technique proposed for automatic methods to detect voluntary eyeblinks. However, the EOG method needs an exclusive apparatus, and the skin must be directly equipped with an electrode. For that reason, it is unsuitable for a simple interface. Moreover, there exists the disadvantage of easily mixing in noise from a living body.

On the other hand, the VTR method is suitable for a computer interface because it is a non-contact technique and has great flexibility. However, the process of change is difficult for a typical NTSC video camera to capture, because an eyeblink is a high-speed operation. Therefore, analysis using a high-speed camera has been attempted [5]. Computer interfaces using eyeblinks have been proposed in several papers [6][7][8]. Eyeblink switches were devised by using two different types of hard instructions in these techniques. For example, in [7], a subject closed an eye for a blink time of over 200 ms but, in [8], double-blinking was used for a mouth clicking switch. We aimed for a method to lower the stress of the subject. Eyeblink waveforms exhibit large individual variations, so automatic detection with multiple persons as the target is difficult. However, we have developed an algorithm such as that described below.

ISBN: 978-1-61208-250-9

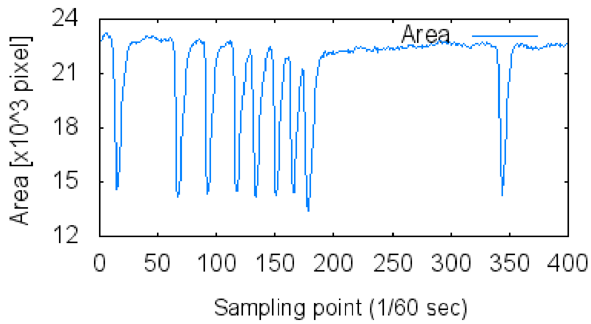


Figure 1. Changes of opening eye aperture area.

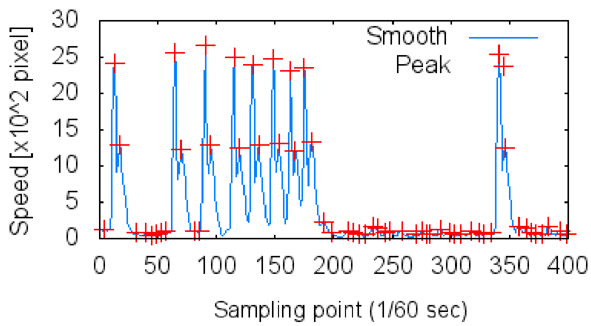


Figure 2. Coordinates of speed maximums.

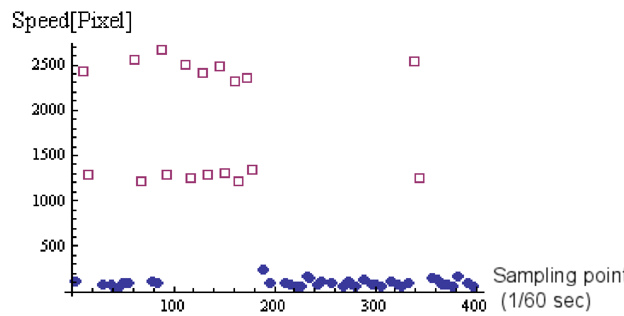


Figure 3. Result of clustering.

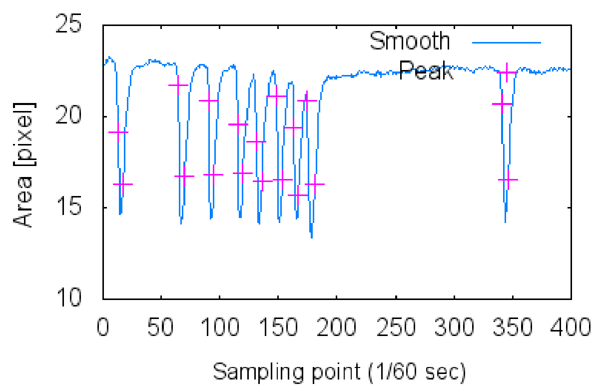


Figure 4. Coordinates of retrieved speed maximums an diagram that records.

The first step is to analyze video images of the near-eye area to obtain the changes in eye aperture area (Fig. 1). The data include the eyeblink waveform. This step incorporates an algorithm used in previous research [9]. This technique shows that we can detect a change in eye aperture area while sampling at 1/60 s by using interlaced NTSC video images divided into field images.

The next step is to differentiate the area data obtained. We smooth the derivative of the aperture area, and then we take the absolute value. The coordinates of the maximums are determined by using a second differentiation. Fig. 2 shows the coordinates of these maximums. Here, the eyeblink operation (namely, the closing and opening phases) corresponds to maximums with large values. However, the result of this step contains much noise. In other words, the maximums due to small changes in the coordinates are from noise. Therefore, we needed to remove the noise by using the *k*-means method for optimal partition clustering. We set the number of clusters to two, because the purpose was separation of the required data part from the noise part. We applied the evaluation function

$$J_b = \sum_{i=1}^{k} \sum_{x \in C_i} \| x - c_i \|^2 . \qquad (1)$$

where J_b represents the objective function, C is the dataset, x is a data value, and c is the cluster centroid. The *k*-means process is repeated until J_b is minimized, and then we terminate the clustering process. Fig. 3 shows the clustering result. Of the clusters obtained, the lower cluster is the noise cluster, and the upper cluster is the required data cluster. We then plotted the coordinates of the retrieved maximums on the diagram that records the changes of eye aperture area (Fig. 4).

There are times when a strange movement of the eyes is noted in the middle of a blink and maximums are recorded in more than three places. Then, if a clustering process is performed, it will not be able to remove superfluous points and will not be able to pick up the eyeblink waveform normally. Hence, we need to remove the superfluous points. We solve this problem by removing as noise the points that were recorded later, if the recording was continuous over a short period of time.

As a result of these processes, we obtain maximums in the middle of the eyeblink waveform. We can distinguish closing and opening by evaluating the derivative of the eye aperture area at that point. In other words, if the derivative at the point is negative, the point exists in the closing phase. Conversely, if the derivative is positive, the point exists in the opening phase. We can determine the eyeblink starting point and the eye closing point based on this. Moreover, we can determine the duration of a single eyeblink waveform.

 ISBN: 978-1-61208-250-9

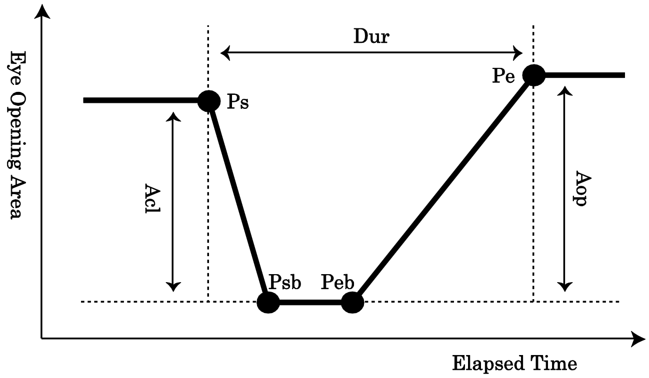


Figure 5. Shape feature parameters of eyeblink.

III. SHAPE FEATURE PARAMETERS OF VOLUNTARY EYEBLINK

Blinking can be categorized as voluntary, reflex, or spontaneous [10]. A voluntary eyeblink occurs consciously and is performed at the order of the experimenter. A reflex eyeblink occurs due to external stimulation such as photic stimulation or auditory stimulation. A blink that occurs unconsciously is called a spontaneous eyeblink. In this paper, we intend to discriminate between voluntary and spontaneous blinks. Spontaneous eyeblink waveforms have various amplitude patterns and greater individual unevenness than other kinds of eyeblink waveforms. Nevertheless, a voluntary eyeblink causes the eyelids to close almost completely. Moreover, there is relatively little variation in the recorded eyeblink waveforms of the same person. Thus, based on the literature [11], we focused on the parameters of two types of eye blinking. The parameters are the amplitudes in the closing and opening phases as well as the duration. Fig. 5 models a single blink. In this figure, the amplitude of the closing phase (Acl) is measured from the blink start point (Ps) to the eye aperture area minimum (Pmin) between the closing phase end point (Psb) and the opening phase start point (Peb). The eyeblink duration (Dur) is the number of samples from the blink start point (Ps) to the blink end point (Pe).

IV. EXPERIMENTS

A. System Outline

The hardware for our experimental system includes one digital camcorder, which acquires pictures near the eye, and a personal computer, which conducts image analysis and eyeblink waveform analysis. Although the camera can also take high-definition (HD) pictures, the standard-definition (SD) picture was used for the experiment. In addition to ordinary indoor lighting (incandescent lighting), the experiment used LED lighting (LPL: LED VL-300C L26811).

TABLE I. RESULTS OF AUTOMATIC DETECTION OF EYEBLINKS.

	Eyeblinks [number]	Error [number]		Agreement rate [%]
		oversight	*misdetection*	
Subject 1	12	0	0	100.0
Subject 2	4	0	0	100.0
Subject 3	4	2	0	50.0
Subject 4	2	0	0	100.0
Subject 5	4	0	0	100.0
Total	26	2	0	92.3

Incandescent lighting was set when taking the moving images. Then, LED lighting was installed at a distance of about 60 cm from the front of the subject's face. Subjects would sit on a chair at a distance of about 20 cm from the video camcorder, and the back of the head was lightly supported with a device to prevent the head from shaking. The video camera was placed in front of and below the subject`s head. Then, the camera magnified and took pictures around the subject's left eye. The image format is SD video, so the resolution is 720 pixels by 480 lines with 16:9 aspect ratio and the refresh rate is 30 frames per second (NTSC). These experiments were performed on the naked eye, so eye glasses were not allowed during the filming in the experiments.

B. Automatic Detection of Eyeblink Waveform

We conducted an evaluative experiment for the automatic detection of the eyeblink waveform by using the algorithm described in Section II. This experiment was used for preprocessing purposes and to extract the shape feature parameters of the eyeblink waveform. Five people were included in this experiment: four men in their twenties and one man in his thirties.

1) Method: Before the experiment, the subjects were first instructed to "pay attention to the mark that was installed on the upper part of the camcorder" and told "you do not have to resist the unconscious urge to blink." Rest time was provided for about 1 min prior to beginning the experiment. Once the experiment began, moving images were taken for 10 s. Changes in the eye aperture area were obtained from this moving picture by image analysis, and automatic detection of eyeblink waveforms was achieved by using the proposed algorithm.

In this experiment, the obtained waveforms mixed some voluntary eyeblinks with many spontaneous eyeblinks, because the subjects were not instructed to blink. In addition, the instances of automatic detection were validated by comparing the data with those of real blinks observed visually.

TABLE 2. RESULTS OF SHAPE FEATURE PARAMETER EXTRACTION 1 (SPONTANEOUS EYEBLINKS).

	Blinks [number]	Amplitude (Closing) [pixels]	Amplitude (Opening) [pixels]	Duration [ms]
Subject A	19	14589	14964	335.83
Subject B	25	13024	12076	490.67
Subject C	27	15830	15190	356.67
Subject D	3	15562	12635	422.17
Subject E	30	5043	2962	197.67
Subject F	15	10138	10453	556.67
Subject G	3	9397	9352	394.3
Subject H	2	11006	10720	358.3
Subject I	21	11799	10474	506.3
Subject J	16	10039	10318	534.3

TABLE 3. RESULTS OF SHAPE FEATURE PARAMETER EXTRACTION 1 (VOLUNTARY EYEBLINKS).

	Blinks [number]	Amplitude (Closing) [pixels]	Amplitude (Opening) [pixels]	Duration [ms]
Subject A	8	19249	23379	545.8
Subject B	9	14016	13281	1238
Subject C	9	16354	15908	396.1
Subject D	9	16581	15155	762.8
Subject E	9	8712	5320	214.6
Subject F	9	12482	11434	737.0
Subject G	9	12370	12688	474.0
Subject H	9	10459	11045	785.2
Subject I	9	19724	17667	588.8
Subject J	9	13581	12741	498.2

TABLE 4. RESULTS OF SHAPE FEATURE PARAMETER EXTRACTION 2 (SPONTANEOUS EYEBLINKS).

	Blinks [number]	Amplitude (Closing) [pixels]	Amplitude (Opening) [pixels]	Duration [ms]
Subject K	16	12276	12353	346.9
Subject L	13	10077	9909	371.8
Subject M	10	7503	6506	370.0
Subject N	13	7366	5859	421.8
Subject O	19	11775	8085	384.2

TABLE 5. RESULTS OF SHAPE FEATURE PARAMETER EXTRACTION 2 (VOLUNTARY EYEBLINKS).

	Blinks [number]	Amplitude (Closing) [pixels]	Amplitude (Opening) [pixels]	Duration [ms]
Subject K	9	14196	14388	353.7
Subject L	9	12227	12070	381.6
Subject M	9	13044	12396	787.0
Subject N	9	19737	19397	653.7
Subject O	9	12316	11288	807.4

2) Result: Table 1 shows the number of automatic detections and the number of confirmed eyeblinks for each of the subjects. The overall agreement rate of these blink numbers is 92.3%. This result was satisfactory enough for preprocessing. Subject 3 produced a low detection rate; however, a factor in this result is the number of oversights in detection. This factor makes little difference, because the problem can be solved by promoting re-input if the algorithm is implemented as the interface.

C. Shape Feature Parameter Extraction 1 (Experiment 1)

Based on the results of the evaluative experiment of Section IV-B, we performed an experiment to extract the shape feature parameters described in Section III for eyeblink waveforms. Ten people were included in this experiment: nine men and one woman in their twenties. These subjects were all different persons from those in the abovementioned preprocessing experiment.

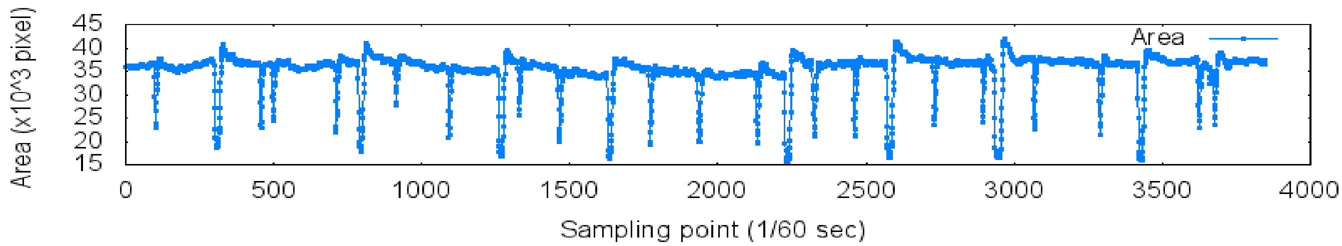


Figure 6. Example of eyeblink waveform.

1) Method: As in the experiment described above, the subjects were instructed before the experiment to “pay attention to the mark that was installed on the upper part of the camcorder” and told “you do not have to resist the unconscious urge to blink.” In addition, the subjects in this experiment were told, “you must always blink when you hear the signal.” This instruction makes it possible to distinguish voluntary eyeblinks from spontaneous eyeblinks. The signal was sounded randomly at intervals of from 4 s to 8 s by using a digital timer. A rest time of about 1 min was provided prior to the task. Once the experiment began, moving images were taken for approximately 1 min.

2) Result: Fig. 6 shows an example of the acquired waveforms (changes in eye aperture area). Moreover, Tables 2 and 3 list the shape feature parameters that were obtained in this experiment. Those for spontaneous eyeblinks are listed in Table 2, and those for voluntary eyeblinks are listed in Table 3. The value of each parameter is the mean for the detected blinks.

D. Shape Feature Parameter Extraction 2 (Experiment 2)

In addition to the experiment described in Section IV-C, an experiment was carried out in another room. Five of the subjects who generated the data presented in this section differed from those who participated in the experiments of Sections IV-B and IV-C (four men and one woman in their twenties).

1) Method: Basically, the experimental method was the same as that described in Section IV-C-1. However, there were two differences. The light of the sun was not evident in the room, and the camcorder was moved to a position directly in front of the subject’s face.

2) Result: Tables 4 and 5 list the shape feature parameters that were obtained in this experiment. Like Tables 2 and 3, these tables each list statistics for spontaneous or voluntary eyeblinks, respectively. The value of each parameter is the mean for detected blinks.

V. DISCUSSION

First, we were able to obtain waveform data regardless of the influence of sunlight, because we compared the data of Experiment 1 with that of Experiment 2 and there was no great difference between the two sets of data.

We discuss the results on the basis of the assumption mentioned above. Fig. 6 shows an example of the acquired waveforms (changes in eye aperture area). In addition, Figs. 7–9 show the patterns of waveform pairs for three types of subject. To enable comparisons between a pair of eyeblink waveform patterns, these plots were normalized using the pixels of the eye aperture area at the first sampling point of the first field image. When we reviewed the average values in Tables 2–5, the shape feature parameters of the voluntary eyeblinks recorded were nearly all greater than those of the spontaneous ones. This was a common trait found with all subjects. Admittedly, the recorded amplitude of the closing phase for Subject H in Experiment 1 was an exception, because so few spontaneous eyeblinks were found during the experiment. However, this voluntary parameter is greater than the spontaneous one if the length of time taken for the experiment is added. Hence, it is difficult for discrimination methods to use the rate of amplitude change because there is an absence of variety in the patterns.

A significant difference existed in the blink durations of the majority of subjects. In particular, Subjects B and H of Experiment 1 had values recorded for voluntary blinks that were twice those recorded for spontaneous blinks. Subject E had a small difference in duration but large differences in the amplitudes of the opening and closing phases. Subject J had a duration for spontaneous blinks that was longer than that for voluntary ones. The reasons are as follows: Subject J was observed to produce many blinks with larger ocular movements than the other subjects. When we accumulated the numbers of blinks, a voluntary eyeblink was counted whenever the subject responded to the signal heard. Consequently, it is thought that among the spontaneous eyeblinks were included some voluntary eyeblinks. Moreover, the change of eye aperture area by ocular movement is a factor that extends the duration.

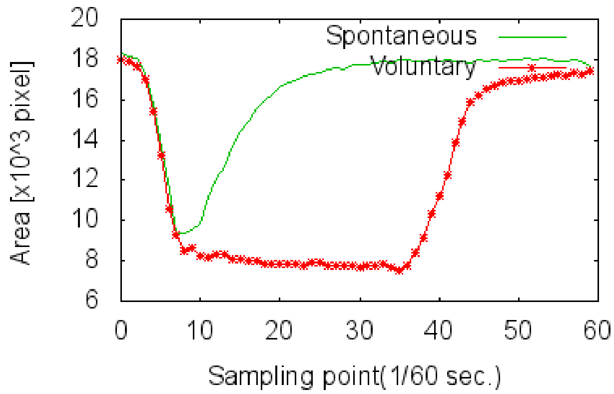


Figure 7. Waveforms recorded for the type of subject with a large difference in duration between voluntary and spontaneous blinks.

Next, we classified the subjects on the basis of three patterns in the shape feature parameters (Figs. 7–9): 1) subjects with a large difference in duration between voluntary and spontaneous blinks (e.g., Subjects B and D), 2) subjects with a large difference in amplitude between voluntary and spontaneous blinks (e.g., Subjects A and I), and 3) subjects with both characteristics mentioned above (e.g., Subject M). There are meaningful differences between these patterns in parameters that reveal each large difference.

Hence, we expect that automatic discrimination of voluntary eyeblinks can be realized by using a machine that has learned to weight the parameters that exhibit large differences. In addition, we think that classifying patterns would become more accurate if the velocities of the eye opening and closing phases were included among the parameters.

VI. CONCLUSION

We examined automatic discrimination of voluntary and spontaneous eyeblinks, which is a problem that has been awaiting solutions in various kinds of blink research. So far, eyeblink detections have needed to use the EOG method or a high-speed camera. However, this paper has proposed a detection algorithm by using ordinary interlaced NTSC video images divided into field images. This algorithm can automatically pick up every eyeblink waveform from the changes in the eye aperture area, which were obtained from an SD moving image by means of image analysis. In our experiments to evaluate this algorithm, the average rate of detection was 92.3%.

Further, we conducted experiments on shape feature parameter extraction using an audio signal and our proposed algorithm. In this experiment, we instructed the subjects to blink voluntarily in order to yield an eyeblink waveform of both voluntary and spontaneous blinks. We instructed the subjects by saying, "you must always blink when you hear a signal."

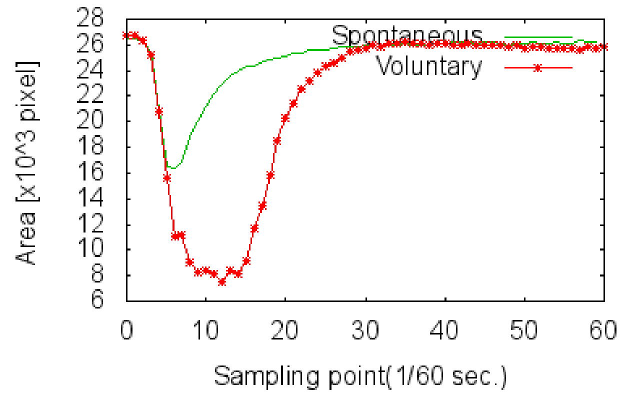


Figure 8. Waveforms recorded for the type of subject with a large difference in amplitude between voluntary and spontaneous blinks.

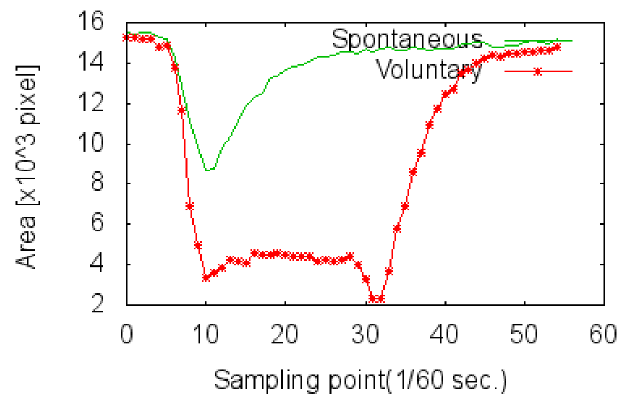


Figure 9. Waveforms recorded for the type of subject with both characteristics mentioned above.

After that, we analyzed the tendencies of the shape feature parameters, and three facts became known. First, the parameters of voluntary eyeblinks are greater than those of spontaneous ones. Second, a large difference in the blink duration is usually observed, even apart from these parameters. Third, a different tendency for large feature parameters is observed with every subject. However, we knew that the rate of amplitude change would be difficult to use as a shape feature parameter.

In particular, we newly observed subjects who exhibited both of the different patterns found in Experiment 2. In other words, we understood that we could classify subjects using three types; namely, 1) subjects with a large difference in duration between voluntary and spontaneous blinks, 2) subjects with a large difference in amplitude between voluntary and spontaneous blinks, and 3) subjects with both characteristics mentioned above.

From the above results, we showed the possibility of discriminating voluntary eyeblinks from spontaneous eyeblinks by using shape feature parameters such as the eyeblink duration and the amplitudes of the opening and closing phases. Moreover, we confirmed the ability to process these parameters in a uniform manner, because we could obtain some tendencies from the parameters even

 ISBN: 978-1-61208-250-9

when the environmental conditions, such as the lighting or the subjects, were changed.

We now wish not only to classify more accurately using other parameters such as velocity but also to classify subjects automatically using the three patterns. We are aiming for real-time discrimination of voluntary eyeblinks and spontaneous eyeblinks.

REFERENCE

[1] J. A. Stern, L. C. Walrath, and R. Goldstein, "The endogenous eyeblink," Psychophysiology, vol. 21, no. 1, Jan. 1984, pp. 22–33.

[2] K. Tanabe, "Eyeblink activity during identification of Katakana characters viewed through a restricted visual field," IEICE Trans. Fundam. Electron. Commun. Comput. Sci., vol. E87-A, no. 8, 2004, pp. 2189–2191.

[3] J. Hori, K. Sakano, and Y. Saitoh, "Development of communication supporting device controlled by eye movements and voluntary eye blink," Proc. IEEE Conf. Eng. Med. Biol. Soc., vol. 6, 2004, pp. 4302–4305.

[4] O. Tetuya, K. Hironori, and K. Masashi, "Development of an input operation of the communication tool using voluntary eye blink," Papers of Technical Meeting on Medical and Biological Engineering, IEE Japan, vol. 6, 2006, pp. 1–4.

[5] K. Ohzeki and B. Ryo, "Video analysis for detecting eye blinking using a high-speed camera," 40th Asilomar Conf. Signals, Systems and Computers (ACSSC '06), 2006, pp. 1081–1085.

[6] K. Grauman, M. Betke, J. Gips, and G. R. Bradski, "Communication via eye blinks — detection and duration analysis in real time," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR 2001), vol. 1, 2001, pp. 1010–1017.

[7] A. Krolak and P. Strumillo, "Vision-based eye blink monitoring system for human-computer interfacing," Proc. Conf. Human System Interactions, 2008, pp. 994–998.

[8] D. O. Gorodnichy, "Second order change detection, and its application to blink-controlled perceptual Interfaces," Proc. IASTED Conf. Visualization, Imaging, and Image Processing, 2003, pp.140–145.

[9] A. Kiyohiko, O. Shoichi, and O. Minoru, "Automatic method for measuring eye blinks using split-interlaced images," Proc. Conf. Human-Computer Interactions, Part I (HCII 2009), LNCS 5610, 2009, pp. 3–11.

[10] F. VanderWerf, P. Brassinga, D. Reits, M. Aramideh, and B. Ongerboer de Visser, "Eyelid movements: Behavioral studies of blinking in humans under different conditions," J. Neurophysiol., vol. 89, 2003, pp. 2784–2796.

[11] B. Matteo, A. Rocco, B. Gregori., D. Belvisi, D. Ottaviani, et al., "Voluntary, spontaneous and reflex blinking in patients with clinically probable progressive supranuclear palsy," Brain, vol. 132, no. 2, 2009, pp. 502–510.

Cursor Control Trace

Another look into eye-gaze, hand, and eye-hand pointing techniques

Ricardo Sol
Exact Sciences and Engineering Centre
University of Madeira
9000-390 Madeira, Portugal
rsol@alumni.carnegiemellon.edu

Mon-Chu Chen
M-ITI, Madeira Interactive Technology Institute
University of Madeira
9000-390 Madeira, Portugal
monchu@uma.pt

José Carlos Marques
Exact Sciences and Engineering Centre
University of Madeira
9000-390 Madeira, Portugal
marques@uma.pt

***Abstract*—We analyzed cursor control trace with respect to three cursor control methods: eye-gaze, hand, and eye-hand. First, we look into the mechanism that allows users to control cursor positions by hands and/or eyes. Second, we conducted an experiment in which subjects perform a searching, pointing and selection task in three different conditions (eye-only, hand-only, and eye-hand). Third, we further studied the cursor trace and analyzed the moments when users switch between eye-gaze-control and hand-control. Although different from a simpler Fitt's pointing task, our results mostly corroborate previous work. In addition, the cursor traces analysis further shows why eye-hand is more efficient, and how users progress from an inefficient pointing behavior to an optimal one.**

Keywords; pointing; accuracy; gaze; eye; tracking

I. Introduction

Hands move slower than eyes. Therefore having eye tracking instead of manual input as an interaction mechanism in digital devices can make interactions faster. However, the eye trackers developed so far have numerous drawbacks. First, current eye trackers are not precise enough to accomplish common tasks on applications or websites. Next, these systems have a delay in processing gaze position and calibration. A third problem is the 'Midas touch', a problem identified by Jacob in 1991, that occurs because it is challenging to distinguish a selection task (for example, a click) from the search task, using a purely eye-gaze approach [2]. Eye pointing is also typically associated with fixation (dwell) but this action is not stable and shows jitter.

In order to address these issues, in 1999 Zhai et al. presented Mouse and Gaze Input Cascaded (MAGIC), a technique that uses a combination of gaze and manual input [8]. First, the user uses gaze to dynamically redefine the position of the cursor. After the cursor position is redefined the user then makes a small manual input action to select the target.

The advantages of MAGIC are that it reduces manual stress and fatigue. In addition, MAGIC pointing is faster than just manual pointing. It must be noted however, that because the cursor movement is faster while controlled by the eye, the user may perceive all action to be faster even when it is not. For example the manual selection time might take longer, so overall the tasks might not be accomplished faster although the user may perceive them to be so.

MAGIC has two approaches. In the first approach, referred to as the liberal approach, the cursor moves directly over (in front of) the target that the users looks at. In the second approach, the conservative approach, the cursor moves to the boundaries of the target.

Zhai et al conducted experimental validations of the MAGIC technique. These experiments were conducted with a miniature isometric pointing stick [8]. The experimental task was basically a Fitts' pointing task. The factors manipulated in the experiment were: target size, target distance, and direction. Each subject performed the task with 3 techniques: no gaze, liberal, and conservative. The liberal technique was found to be faster and preferred by users.

While the MAGIC technique addresses some of the problems of preceding gaze input mechanisms, it is not without its drawbacks. For example, with MAGIC's liberal technique the cursor movement can be overactive, which could be distracting when reading. With MAGIC's conservative technique the user might tend to activate the cursor manually.

Our broad research goal is to improve input techniques. In this work, our contribution adds to previous work by analyzing the trace before, during, and after that manual activation, and we try to find patterns in these traces.

II. Related Work

In 2000 Salvucci and Anderson presented their intelligent gaze-added interfaces [6]. They addressed accuracy problems that we also face. In their work, any target positioned where the users' eye gaze is, is a highlighted target. Then a gaze key gives the user the chance to trigger the action. The system uses a probabilistic algorithm to try to guess the targets the user will look at.

In 2003 McGuffin and Balakrishnan showed that expanding targets facilitates the pointing task [4]. Their results show that working with expanding targets can be accurately modeled by Fitts' law. They have also shown that targets that expand just as the user is about to reach them can be acquired approximately as fast as targets that are always in an expanded state. They specifically found strong evidence that the user performance is consistently aided by the target expansion. Similar to how we hypothesize for our work, they found that the performance is dependent on the final target size.

In 2005 Miniotas and Spakov used an expansion of targets visible to the users [5]. To facilitate pointing they used dynamic target expansion for fixing the calibration of the eye tracker. This technique has 91% accuracy, a result not expected in our work. The drawback of this technique is an increase in selection time.

In 2005 Ashmore and Duchowsky refined a fisheye lens to support eye pointing [1]. They simply hid the lens during visual search and obtained improvements in speed and accuracy. Fisheye interaction was evaluated by a visual search, and a Fitts' pointing. Unlike MAGIC pointing, where the cursor was rapidly moved to the vicinity of one's gaze prior to mouse movement, here the lens is directly slaved to gaze position. Important to our work is their finding that in combined tasks it is impossible to distinguish the precise amount of time that search consumed prior to selection.

In 2007 Kumar et. al presented EyePoint. EyePoint uses expansions of interactive targets, and uses a key for input [3]. When the key is pressed the gaze area is enlarged. When the key is released the selections are made according to where the eye gaze is. Similar as we hypothesize for our work, they found that it is possible to divide appropriate interaction techniques that use gaze without overloading the visual channel.

III. METHOD

A. Participants

The experiment was performed on 10 participants with normal or corrected-to-normal vision. All participants were regular computer and mouse users. All participants had used track pads.

B. Tasks

Three conditions (eye-only, hand-only, and eye-hand) were presented that required searching and selecting one target. In the hand-only condition the participant used the track-pad to move the cursor until it was over the target. In the eye-only condition the participant used eye-gaze to move the cursor until it was over the target. In the eye-hand condition the participant used eye-gaze and the track-pad to move the cursor until it was over the target. The target consisted of a white vertical rectangle with a plus signal in its center. This rectangle could alternate in size (2% or 4% of the screen width), and distance to the center along the screen's horizontal central axis (45% or 90% of half of the screen width), giving a total of four different possibilities. The target also alternated between sides of the screen. Each trial had one distractor with the same characteristics as the target without the plus signal. All trails started with the cursor located at the center of the screen as in Fig. 1 and Fig. 2. All trails finished with the participants pressing the space key while the cursor was over the target. Each trial had a maximum amount of time, 5 seconds, and if this expired without pressing space while on the target, the task failed.

C. Apparatus

Participants used a track-pad to complete the task (Apple Magic Trackpad).

The experimental computer ran Mac OS X version 10.7.4 and was connected to Tobii TX300 Eye-tracker. The system allowed for unconstrained head motion by seating participants approximately 65 cm (adjusted by Tobii studio running on a PC) in front of the Eye-tracker 23 inch screen (resolution 1920 x 1080 pixel). With this system the sample rate was 300 Hz.

Figure 1. Initial display of a trial that has the target (right side) with the biggest size at the farthest distance.

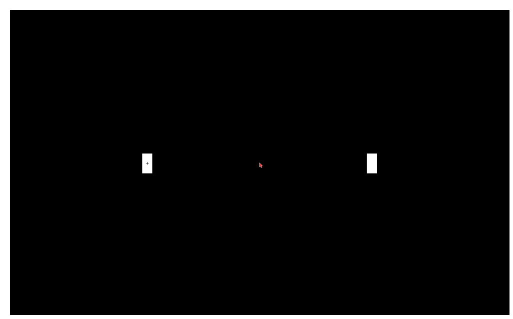


Figure 2. Initial display of a trial that has the target with the smallest size at the closest distance.

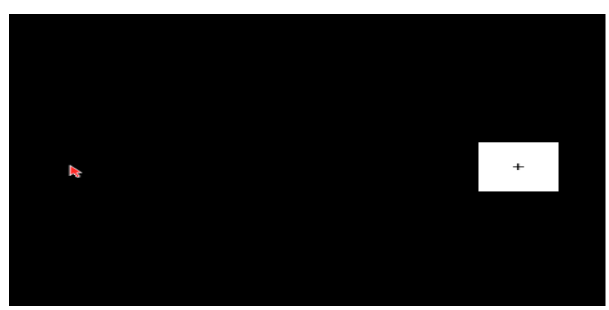


Figure 3. A zoom in of the initial display of a trial that has the target with the biggest size at the closest distance.

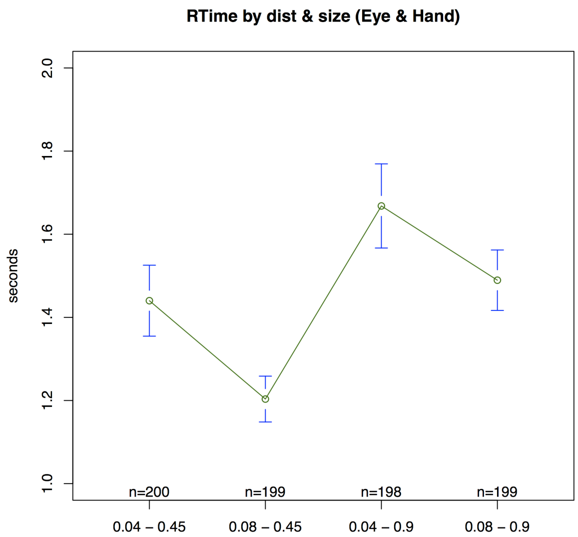


Figure 4. One trial of one task presented to the users. In this case the target is small and is farther from the cursor's original point.

D. *Procedure*

The user was first asked to sit in a non-mobile chair. Then the distance of the eyes to the screen was adjusted and the user was asked not to make large movements with his head and torso. After this, it was necessary to calibrate the user's gaze to the system. This calibration was made by having the user looking each time for 3 seconds to a dot that appeared in 9 different pre-selected positions of the screen. Before data collection, the conditions and task were explained, and unlimited time to practice with the 3 conditions was given. The data collection started with the conditions ordered randomly. Each participant then carried out 80 trials for each condition. Each experimental session lasted around 45 minutes including pauses between conditions.

IV. RESULTS

Our results show that Eye-only is significantly faster than Hand-only solely when the target has the largest size and is at the farthest distance. The variance of the response time is always higher in Eye-only. Eye-only has a hit on target rate of 72%. That is, participants successfully accomplished the task within the limit of 5 seconds in only 72% of the trials in the Eye-only condition. These hits on target were lower in the smallest size and farthest distance condition.

The results also show that Eye-with-Hand is the significantly fastest solely when the target has the largest size and is at the closest distance. The variance of the response time in Eye-with-Hand is higher than in Hand-only and lower than in Eye-only, as Figure 4 shows. Eye-with-Hand has a hit on target rate of 99.50%. This is similar to Hand-only that had a hit on target rate of 99.75%.

Our results also show the trace of each trial for Eye-with-Hand. In Figure 5 we can see an early trial trace. This user is not yet sufficiently familiarized with eye-gaze control. The target is situated in the right side of the screen, that is 0.9 on the vertical axis. The user first briefly looked to the left and immediately changed direction to the right and changed it again to the left all way. Then the user changed control to the hand bringing the cursor to slightly right of the center. The users' finger then reached the right extremity of the track pad, so the user released the finger, passing automatically to the control of the eyes. The user subsequently acquired control with the track pad bringing the cursor to the extreme right of the screen. Then the users' finger reached the right extremity of the track pad, so the user released the finger passing automatically to the control of the eyes. The user then looked to the center and had to repeat the process twice, failing to press space while the cursor was on top of the target, and not completing the task in the defined time.

In the trial in Figure 6, the user started looking to the left and then looked to the right. 1 second after the beginning, the user started controlling the cursor with the hand and moved the cursor to the target and at 2.1 seconds pressed space.

In Figure 7, the user started looking to the left and then looked further to the left, then looked back to center, and back to the left again. 1.4 seconds after the beginning the user started controlling the cursor with the hand, moved the cursor to the target and at 2.4 seconds pressed space.

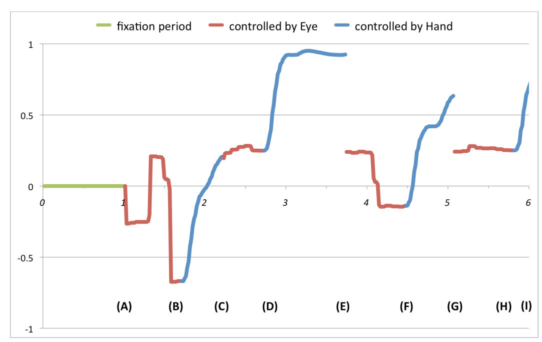


Figure 5. Eye control and Hand control shown over Time as a function of Distance. Early trial.

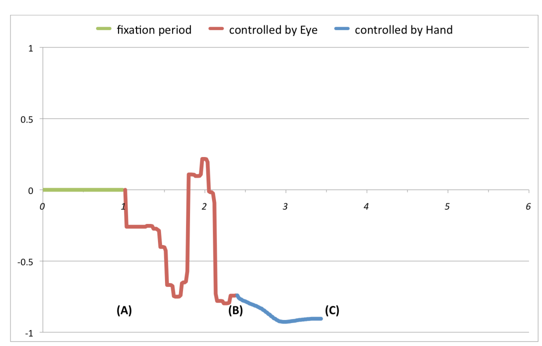


Figure 6. Eye and Hand control over Time as a function of Distance.

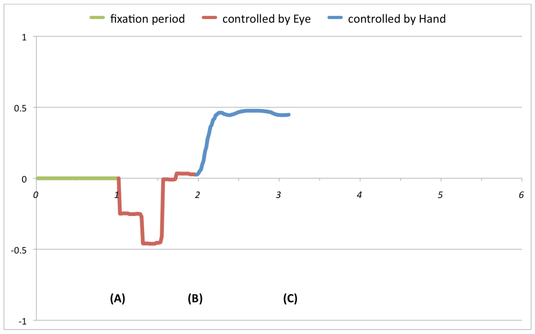


Figure 7. Eye control and Hand control shown over Time as a function of Distance.

V. CONCLUSION & FUTURE WORK

We show support for MAGIC by stating with confidence that in at least in one condition, the Eye-with-Hand condition, the performance was faster. An explanation for this could be that this system had a more optimized transfer function for the movements. Another factor contributing to this could be the more optimized parameters in the gaze system's filter. In addition, we used a more recent eye-tracker as compared with those used in previous work and is therefore expected to be more accurate.

Eye-gaze control is a novelty for users. Users' lack of experience with eye-gaze can introduce delays in the control action. It is expected that with the dissemination of eye-tracker these delays can diminish. However, controlling the cursor with Eye-gaze can give the user a sense of speed.

The results help to further understand how people progress to master controlling cursor using the combination of the eye-gaze and mouse.

We foresee that having a relatively short constant time for hand control will decrease the time taken to reach the target. More distance will be traveled during the fastest control, which is the control by the eye. We foresee that both, eye and hand control conditions are governed by Fitt's Law. We intend to perform more experiments in order to confirm this.

Some users considered having the cursor in the gaze position distracting. While in MAGIC's conservative approach the cursor stays in the boundaries of the target, this can still be a distraction. One potential solution for this problem might be to decrease the cursor's opacity.

In a pointing task, having the cursor in the eye-gaze might not be so distracting. In a selecting task this distraction can be higher. However, it is during reading that we expect to find the cursor in the eye-gaze to be most distracting. In future work, we intend to perform further experiments in order to confirm these hypotheses.

ACKNOWLEDGMENTS

The authors acknowledge the financial support of the Portuguese governmental agencies ADI – Agencia da Inovacao, and FCT – Fundacao para a Ciencia e Tecnologia through PEst-C/CTM/LA0025/2011 and project PTDC/EQU-EPR/114685/2009.

REFERENCES

[1] Ashmore, M., Duchowski, A. T., and Shoemaker, G. Efficient Eye Pointing with a Fisheye Lens. In Proceedings of Graphics interface 2005. ACM International Conference Proceeding Series, vol. 112, 203 – 210, 2005.

[2] Jacob, R. J. K. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. In Proc. CHI '90, pages 11-18. ACM Press, 1990.

[3] Kumar, M., Paepcke, A., and Winograd, T. EyePoint: Practical Pointing and Selection Using Gaze and Keyboard. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '07. ACM Press 421 – 430, 2007.

[4] McGuffin, M. and Balakrishnan, R. Acquisition of Expanding Targets. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '02. ACM Press 57 – 64, 2002.

[5] Miniotas, D., Špakov, O., and MacKenzie, I.S. Eye Gaze Interaction with Expanding Targets. In CHI '04 Extended Abstracts on Human Factors in Computing Systems. CHI '04. ACM Press 1255 – 1258, 2005.

[6] Salvucci, D. D., and Anderson, J. R. Intelligent Gaze-Added Interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '00. ACM Press 273 – 280, 2000.

[7] Spakov, O., and Miniotas, D. Gaze-Based Selection of Standard-Size Menu Items. In Proceedings of the 7th international Conference on Multimodal interfaces. ICMI '05. ACM Press 124 – 128, 2005.

[8] Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (MAGIC) pointing. In Proc. of CHI '99, 246–253. ACM, 1999.

[9] Zhai, S., Conversy, S., Beaudouin-Lafon, M., Guiard, Y. Human On-line Response to Target Expansion. In Proc. of the SIGCHI, 177 – 184. ACM Press , 2003.

A Hybrid Tracking Solution to Enhance Natural Interaction in Marker-based Augmented Reality Applications

Rafael Radkowski, James Oliver
Virtual Reality Applications Center
Iowa State University
Ames, USA
rafael | oliver@iastate.edu

***Abstract*—In this paper a method for enhanced natural interaction in Augmented Reality (AR) applications is presented. AR applications are interactive applications, designed to facilitate the handling of virtual objects, represented by a physical proxy object. Ideally, interaction should be natural, in that the user should not notice its technical realization. Interaction capability relies on tracking technology, which enables spatial registration of virtual objects in 3D. Markers are a common solution for this. However, the marker must stay in line of sight of a video camera. In highly interactive applications, the user's hands regularly cover the markers. Thus, the virtual object disappears. This paper describes a hybrid tracking solution, which incorporates marker-based tracking and optical flow-based tracking. The optical flow-based tracking supports the marker-based tracking: it seeks corners of the marker to keep track of them. If no markers are visible, the optical flow tracking extrapolates the position of the object to track. Thus, the virtual object remains visible. A prototype implementation and example application show the feasibility of the solution.**

Keywords- augmented reality; hybrid tracking; interaction

I. INTRODUCTION

Augmented Reality (AR) technology is a type of human-computer interaction that superimposes the natural visual perception of a human user with computer-generated information (i.e., 3D models, annotation, and text) [1]. AR presents this information in a context-sensitive way that is appropriate for a specific task, and typically, relative to the users physical location. Special viewing devices are necessary to use AR. A common viewing device is the so-called head mounted display (HMD), a device similar to eyeglasses that use small displays instead of lenses.

AR applications are designed to facilitate natural interaction. In general, an interface is described as natural when its technical realization is effectively veiled from the user [2]; i.e., the user should not consider how the interaction is working.

Spatial tracking is a key factor of AR and the interaction within an AR application. The term tracking denotes the continuous determination of the position and the orientation of a physical item (i.e., the user, a video camera, etc.). Marker-based tracking is a commonly used tracking technique in the field of AR. It is based on fiducial markers, which incorporate a defined geometry for position and orientation estimation and a unique pattern for identification. A video camera captures an image of a maker and a computer vision algorithm locates and identifies the marker in the image and estimates the spatial relationship between the marker and camera. The pose of that marker is thus known and a 3D model can be rendered and composited with the video image with correct alignment to the user. Usually markers need to be attached on the body of a physical object to track it. Marker-based tracking is popular, because applications are easy to deploy, the technique is convenient to use, and the algorithms are mostly free. The ARToolkit, for instance, is a common tracking system that works this way [3].

However, marker-based tracking technology can hinder the naturalness of the AR user interface. Marker tracking requires a free line of sight between camera and marker [4]. Today solutions suffer performance degradation due to the hands of the user that partially covers the markers. For example, in a highly interactive application, e.g., an assembly training application, the user's hands regularly cover the marker. An assembly training application displays assembly instructions as 3D models and text annotations spatially registered to physical objects. The user needs to grasp these objects in order to assemble them. While doing so the user's hands cover a large area of the scene often occluding the marker from the video camera's field of view. The consequence is, the virtual information disappears. This problem occurs in many highly interactive applications.

To address this problem, this paper describes a hybrid marker-based tracking system. The tracking system uses multiple video cameras. The marker information of all cameras is merged to one so-called marker object model. Thus, the probability of losing the marker is reduced. In addition, the marker and physical object movement is analyzed using optical flow. When no marker is in the line of sight of a video camera, its position and orientation propagation can be estimated using a Kalman Filter-like approach. Thus, all markers can be hidden for a short time.

The paper is structured as follows. In the next section related work is introduced. Section 3 describes the hybrid tracking system, which is the basis for the enhanced natural feature tracking. Section 4 demonstrates the advantages of this approach for enhanced natural interaction. An assembly application is used as an example. The paper closes with a summary and an outlook for future research.

 ISBN: 978-1-61208-250-9

II. Related Work

The related work addresses two areas: fiducial marker tracking systems and hybrid tracking systems. The first subsection presents existing maker based solutions. The second subsection introduces approaches that increase tracking stability and improve the interaction capabilities. It concludes with a summary that explains the gap.

A. Fiducal marker tracking

Kato et al. [4] presented the ARToolKit marker tracking, which is the basis for this work. The ARToolkit uses template markers that consist of a black border with a unique image embedded. The ARToolkit uses the image to identify the marker and the black border to calculate the position and orientation of the marker with respect to a video camera. It is easy to use and an application is easy to deploy. However, the tracking quality depends on the lighting conditions, and tracking is impossible when a marker is partially covered.

Fiala [5] introduced ARTag, a fiducial marker tracking system that uses and id-enhanced marker. An id marker uses a binary pattern that represents a digital code. The ARTag is more robust than the ARToolkit: a marker can be partially covered without loss of tracking.

Wagner et al. [6] introduced a grid dot marker. They draw a grid of dots on a common map to combine a fiducial marker system (grid of dots) with a feature-based marker system. Thus, the marker can also be partially covered.

Wagner et al. [7] also developed the ARToolkit Plus. It is an enhancement of Kato's ARToolkit that is specialized for mobile devices.

Naimark et al. [8] developed data matrix marker. That is a fiducial marker tracking system with a 2D barcode surrounded by a black frame. The barcode incorporates a set of black and white areas. The pattern of black and white corresponds to a marker ID.

Recently, Uchiyama et al. [9] presented random dot markers. A random dot marker is a set of distributed points on a limited area. This random set of points acts as a template that facilitates the identification of the marker.

In summary, fiducial marker tracking systems work well and are widely used in academic research. However, they have different limitations. For instance, the robustness of the tracking system depends on the image quality and the light conditions in the surrounding environment. If the tracking system retrieves a video image that contains noise, it can cause jitter of the calculated pose. In addition, the tracking algorithm utilizes computer vision, whose feasibility depends on the lighting conditions. If the image processing does not comply with the lighting conditions, it can also causes jitter.

B. Hybrid Tracking

Hybrid tracking systems are developed to address the limitations of a single tracking technology. Usually they merge the tracking data of two or more tracking systems to facilitate robust and jitter-free tracking of a physical object.

Seo et al. [10] developed a hybrid tracking system that solves jitter and occlusion problems of common fiducial markers. They use a fiducial tracking system and enhance the tracking with a corner tracking system. They use Kanade-Lucas-Tomasi Feature Tracker to detect and track additional key points of a marker. This data is merged with the pose calculated by the marker tracking system to facilitate a robust tracking.

Marimon et al. [11] developed a hybrid tracking system that incorporates a marker tracking system and a particle filter-based tracking system. The particle filter searches for feature points of the marker and estimates a 3D pose. The feature points are described as a 2D back-projection of the 3D corners.

The tracking systems presented in [10] and [11] work similar to the approach taken in this research. However, both systems require a corner model of the marker, which facilitates a continuous tracking of the marker. In addition, the systems have been tested with one fiducial marker only. Thus, it is unclear, if they work with multiple markers. Finally, they have not been tested in a realistic application scenario like virtual assembly training.

Piekarski et al. [12] combined a marker tracking system with a GPS (global positioning system) tracking system to facilitate indoor and outdoor tracking with a single system. They attached large-sized markers on the walls and the ceiling of a room and tracked them with multiple cameras. The estimated pose is merged with the GPS position. However, their tracking system facilitates indoor tracking and is not designed to track objects with which a user interacts.

Kalkusch et al. [13] incorporates marker tracking and inertial tracking to facilitate robust indoor tracking. Their system requires markers on the wall of a building to track the position of the user. The inertial tracking acts as a backup system for the marker tracking. The system is also not designed to track objects.

Yang et al. [14] proposed a camera tracking approach that combines inertial sensor tracking and marker-based tracking. They use the data retrieved from the inertial sensor to reduce jitter. Yang et al. [15] combines a marker-based tracking system and a feature-based tracking system to realize a robust tracking for their AR book. They use small-sized marker (2cm x 1.3cm) to identify the pages of the book and combine it with a random tree, a machine learning computer vision method to describe and detect feature points. The pose retrieved from the marker tracking is used as the initial position for the feature tracking. Thus, enhancing accuracy.

Fischer et al. [16] presents a hybrid tracking approach that incorporates tracking data from an infrared tracking system and a computer vision-based tracking system. The infrared tracking system tracks rigid objects. A computer vision algorithm tracks distinct points on the subject to track. The data of both systems is merged to reduce noise.

The research summarized above is only part of all relevant works. Further approaches can be found in [17], [18], and [19]. In general, most of the approaches are similar: they incorporate two or more tracking technologies to address the limitations of a single technology.

However, the hybrid tracking systems do not generally consider the interaction of the user. If an object to be tracked must also be grasped by the user, markers are occasionally

 ISBN: 978-1-61208-250-9

covered. Thus, the marker tracking is limited. Robust tracking in this scenario remains a challenge.

In addition, the tracking systems presented generally use two equivalent tracking systems and merge data from both. In contrast, in the approach taken in this research, only a backup system is necessary that becomes active when the main tracking system is not able to track the fiducial marker. The advantage of this approach is robustness and computational efficiency.

III. HYBRID TRACKING

This section presents the hybrid tracking. First, an overview of the tracking setup and the tracking process is presented. Secondly, the hybrid tracking approach is described in detail, including the mathematical foundations and logic during tracking.

A. Overview

Figure 1 shows an overview of the hardware setup for the hybrid tracking solution. This implementation employs a monitor-based AR application setup in which the resulting superimposed video stream is shown on a monitor, with screen oriented face-to-face with the user. The system is designed for a single user. However, since it is a monitor-based AR application more persons can see the output video image.

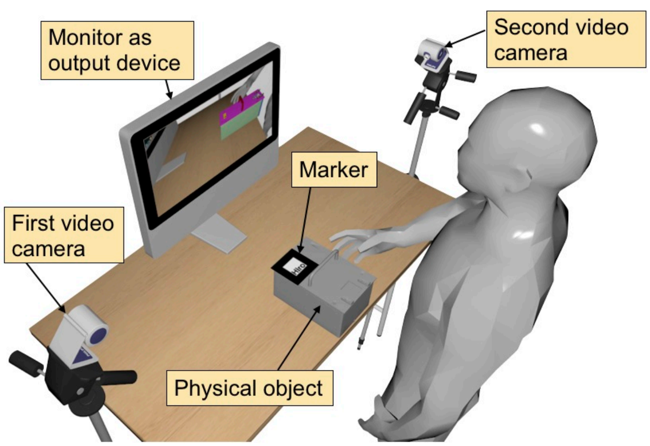


Figure 1. The hardware setup for the hybrid tracking solution.

The user works with a physical object that is located on the table-top work area. It is the object to be tracked. Therefore, one or more fiducial markers are attached on the body of this object. Two cameras are used to retrieve video images for tracking. The cameras must be arranged at different positions around the working area. After calibration, they must remain at their position and cannot be moved. The working area is roughly 1m in square and depends on the resolution of the camera and the size of the marker. This setup is a common setup for industrial applications like virtual assembly training or maintenance training. It is inexpensive, works with common computer hardware, and is easy to deploy.

The approach taken for hybrid tracking in this research is to use the marker tracking as the primary source for tracking and to back it up with optical flow tracking. It works similar to a Kalman Filter with measurement and prediction components.

Figure 2 shows an overview of the hybrid tracking method. The starting point are two images that are retrieved from the video cameras. The left side (green boxes) of the figure depicts the common marker tracking that is extended by a position estimation function. Using a Kalman Filter analogy, this part of the system acts as the measurement. The right side (blue boxes) of the image shows the optical flow tracking. It estimates the movement of an object. Thus, it can be considered analogous to the prediction part of a Kalman Filter. However, it is an extrapolation only. The result is a matrix in homogenous coordinates that describes the pose of the tracked object and is used as transformation data for virtual objects.

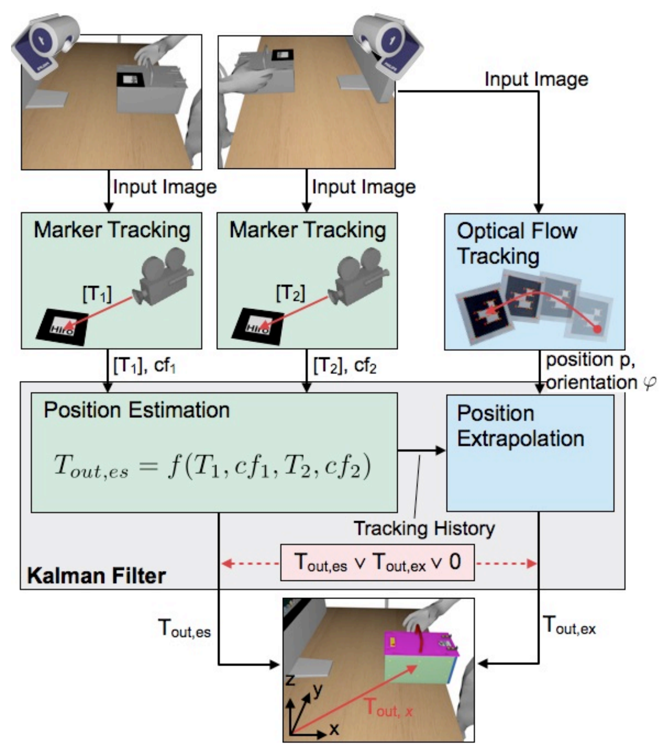


Figure 2. Overview of the hybrid tracking method.

The marker tracking module tracks the fiducial markers that are attached on the body of a physical object. In contrast to the common approaches, multiple cameras are used to track markers. Each module is related to one camera. A module identifies and tracks the markers separately, without information about the other tracking module. It provides a matrix in homogenous coordinates that keeps the position and orientation information of the marker with respect to a global coordinate system, regardless of whether a single marker or a multi marker (a set of multiple single markers that act similar to a single marker) is used. In addition, the module provides a confidence value c for each single marker. This confidence value describes how certain the

 ISBN: 978-1-61208-250-9

module is that the provided information belongs to a marker [3].

The position estimation module gets the information of both marker tracking modules and estimates a position and orientation of the marker (and the tracked object) with respect to a global coordinate system. Figure 2 shows two marker tracking modules, but theoretically, the position estimation facilitates the estimation on the basis of an arbitrary number of tracking modules. The estimated position is denoted as $T_{out,es}$.

The optical flowtracking module tracks corners of the markers. The movement of each corner is tracked frame by frame. The tracking data is merged, and the propagation of the position p and orientation φ is calculated.

The position extrapolation module uses the data of the optical flow tracking to extrapolate the locomotion of the physical object position and orientation. In addition, it gets access to a pose history of the markers. The outgoing data is denoted as $T_{out,ex}$,

Following a Kalman Filter-like approach, the marker tracking data $T_{out,es}$ is used when at least one marker tracking module is able to track a marker. The extrapolated position and orientation data $T_{out,ex}$ is used when no marker tracking module provides tracking information. Finally, the data provides the transformation information $T_{out,x}$ (with x denoting either *ex* or *es*) for spatial registration of a 3D model.

In the following, the marker tracking module and its pose estimation is explained. Then the optical flow tracking and the pose extrapolation is described. In the description the tracking is described for one marker only in order to simplify the presentation. The method can track more than one marker or multi marker.

B. Marker Tracking

The objective of the marker tracking modules is to estimate the position and orientation of a single marker or multi marker with respect to a global coordinate system. Figure 3 shows the relationships between marker coordinates, camera coordinates, and the global coordinate system.

The desired position and orientation of the marker is denoted by T_i., where i is the number of the camera. It is the global position, calculated by one tracking module and camera i. It is calculated by:

$$\mathrm{T}_i = T_{i,w->c} + T_{i,c->m} \tag{1}$$

$T_{i,c->m}$ is the relation between the camera and the marker in the camera coordinate system of each camera. The ARToolkit is used for marker tracking [3], which provides the relation $T_{i,c->m}$.

The position and orientation of each camera $T_{i,w->c}$ is calculated during an initial calibration procedure. Therefore, a calibration marker is used. All cameras must see this marker at the same time. Thus, the marker coordinate system can be assumed as the global coordinate system. All cameras must remain at their position after its global position has been specified.

The result of the marker tracking is a matrix with the global pose T_i of each tracking module i, in homogenous coordinates. In addition, a confidence value cf_i is provided. The ARToolkit provides it for each marker.

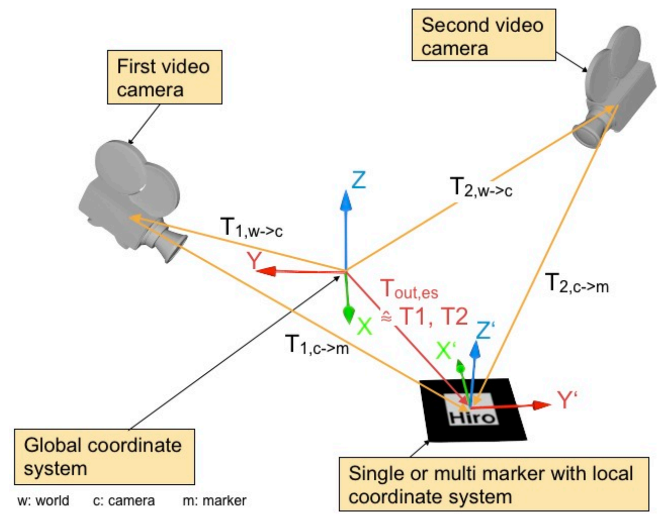


Figure 3. Relationships between the cameras, the marker coordinate system, and the global coordinate system.

C. Pose Estimation

The objective of the pose estimation is to calculate the global position of the marker $T_{out,es}$ with respect to the world coordinate system (Figure 3). Therefore, a linear combination merges all pose data into one resulting value, considering the confidence value of each marker:

$$T_{out,es} = \sum_{i=0}^{n} T_i \frac{c_i}{c_n} \tag{2}$$

with T_i, the value of the single marker and c_i, its confidence value, and n, the amount of used marker tracking modules. The value c_n is the sum of all values c_i:

$$c_n = \sum_{i=0}^{n} c_i \tag{3}$$

Thus, the ratio c_i/c_n rates the quality of the marker tracking.

When a multi marker pattern is used, all single markers provide their own confidence value. In this case the value c_i is the arithmetic average value of all single markers' confidence values.

In addition to the linear combination, two further improvements can be used: a priority of the cameras and a threshold value. The priority of the cameras allows it to specify one camera as a primary tracking camera. This is helpful when it is known in advance that one camera has an uncovered view to the markers most of the time. The priority for a camera, in this case for camera j, can be modeled using the confidence value:

$$c_j \geq c_{threshold} \implies c_i = 0.0 | \forall c_i \ni c_j \tag{4}$$

and

$$c_j < c_{threshold} \implies c_i := max\{c_0 \ldots c_m\} \qquad (5)$$

with $c_{threshold}$, the minimum accepted confidence value, and c_j, the confidence value obtained from camera *j*. This acts as a switch. As long c_j exceeds $c_{threshold}$, the marker data are used. Otherwise, the next best marker data are used for tracking.

The result of this module is a value that describes the pose of the marker $T_{out,es}$ with respect to a global coordinate system.

D. Optical flow tracking

The objective of optical flow tracking is to determine the movement of the physical marker. A commonly used optical flow tracker, the Lukas Kanade-Tracker [20], is adapted for this application. Figure 4 shows the six-step procedure of the optical flow tracking approach. In the first step, a region of interest (ROI) is calculated. The ROI describes an area in the image in which the optical flow tracking can probably find a marker. The ROI is located at the last known position of the marker. Its size correlates to the marker size scaled by the factor 1.3 to ensure that all significant points will be inside the ROI.

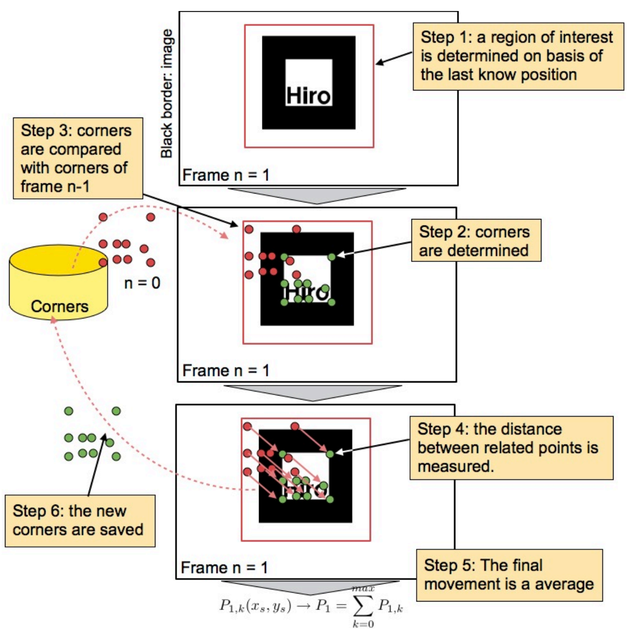


Figure 4. Concept of the optical flow tracking

In the second step, the optical flow tracker searches for corners inside the ROI frame. The Lukas Kanade-Tracker utilizes Shi Tomasi-Corners [21]. Each identified corner *P* is represented as a two-dimensional point $p_k(x_s, y_s)$ in screen coordinates. The corners have to be strong corners according to Harris and Stephens [22]. All corners that are used for tracking have to meet the criteria:

$$P_1 := \{p_k, ROI(p_k) \cup e > e_{threshold} | p_k \in P\} \qquad (6)$$

with ROI, the region of interest function, *e*, the eigenvalue of the corner and $e_{threshold}$, the minimum allowed eigenvalue.

In the third step, the set P_1 is compared with a predefined set of corners. The predefined set of corners P_0 describes the geometry of the pattern. It is defined during an initialization step. For this comparison the Lucas Kanade-optical flow method is used [21]. It provides all points of the set P_1 that are also part of the set P_0:

$$P_{final} := P_1 \cup P_0 \qquad (6)$$

Next the movement of each corresponding point is calculated and summed (Step 5) to one movement vector by:

$$\Delta p = \sum_{k=0}^{max} p_{1,k} - p_{0,k} \qquad (7)$$

with $p_{1,k}$ the corner of the set P_0 and $p_{1,k}$, the corners of set P_1, and *k*, the corner index. The result of this step is Δp, the movement of the marker between two frames in frame coordinates. In addition, the value $\Delta\varphi$, is calculated, which describes the orientation change of the vector Δp.

Finally in step six, the new corner points P_{final} are stored. These steps are continuously repeated.

Figure 5 shows a test application for the optical flow tracking. The main image shows the output image (taken from a video) and contains one marker. A blue box is rendered on top of this marker. The black box on the bottom of the image is a miniaturized window of the main image. The white area shows the region of interest. The window on the lower right corner is a debug window for optical flow tracking. The red dots are the corner points and the red lines depict the movement of the pattern between two frames.

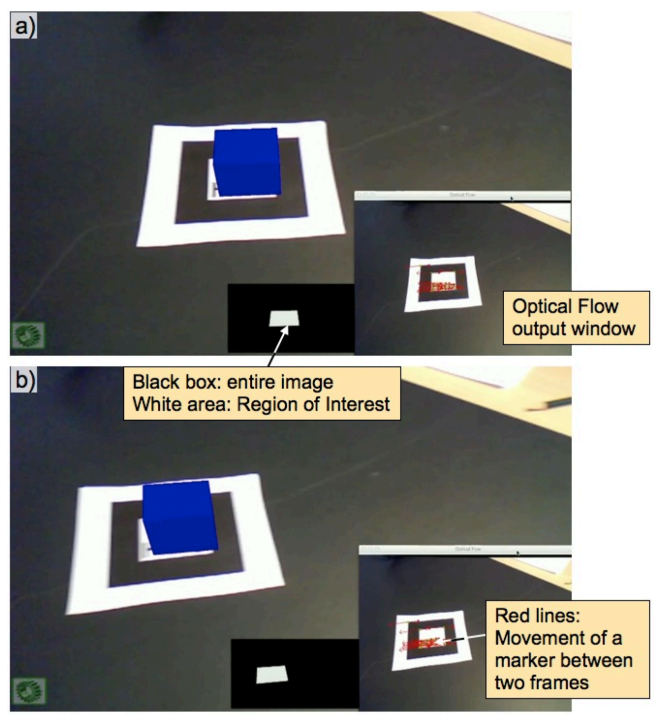


Figure 5. Optical Flow tracking test application

E. Pose Extrapolation

The objective of pose extrapolation is to extrapolate the movement of the recognized pattern and to describe it in a special coordinate system. Therefore, a Kalman Filter-like method is used. In general, a Kalman Filter is a mathematical method that uses measurements of a system observed over time and a prediction model to estimate the current state of a system [23]. Usually, the prediction model is based on a physical model, which describes the behavior of the physical system under observation. The solution developed here works similar to a Kalman Filter in principle. However, the Kalman Filter uses a model to predict the next value. Our approach uses a second measurement and an estimation to predict a next value.

Figure 6 shows an overview of the pose extrapolation process. Input data for this process is the output of the optical flow measurement and the previous, as well as, the current marker tracking data.

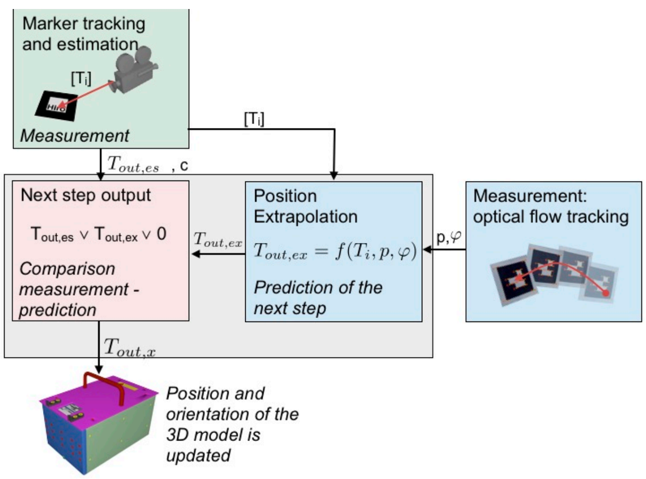


Figure 6. Overview of the position extrapolation

The observed measurement, in the sense of a Kalman Filter, is the marker-based object tracking and estimation process. The prediction process is a position extrapolation, which is based on optical flow tracking. The position $T_{out,ex}$ is extrapolated by a linear approximation of the movement of the physical object:

$$T_{out,ex} = T_{out,es,t-1} + (K \times \Delta p) + (K \times \Delta\varphi) \quad (8)$$

with $T_{out,ex,t-1}$, the previous tracking data, Δp, the movement of the pattern, and $\Delta\varphi$, the orientation of this vector. K is a camera projection matrix, which projects the 2D values of the optical flow from pixel coordinates to 3D coordinates. A common camera projection is used according to the camera calibration according to Tsai [24].

To determine the output for the next step, the marker tracking data and the position extrapolation data are compared. Two error values are used that act as a probability value for both measurements. The probability value for the marker tracking is the confidence value c, which the ARToolkit provides. The probability value c_p for the position extrapolation is a time and distance dependent value that is based on the time of the last update of $T_{out,es}$. and the moving distance Δp:

$$c_p = \begin{cases} 1 & : \quad t = t_0 \\ -\Delta p_s \cdot t + \frac{1}{\Delta p_s} & : \quad t > t_0 \end{cases} \quad (9)$$

with Δp, the movement of the pattern between two frames, t, the current time, and t_0, the time of the last update of the values $T_{out,es}$. Because the position extrapolation is based on the pattern position and orientation, it is assumed that its quality and accuracy decrease with time. Thus, the function is time dependent. In addition, it depends on the movement distance. Short movements are assumed because they can be estimated with a higher accuracy than long movements. This assumption is considered by the vector Δp_s, which is a scaled version of Δp.

The final position T_{out} is determined by:

$$T_{out} = \begin{cases} T_{out,es} & : \quad c \geq c_p \\ T_{out,ex} & : \quad c_p > c \wedge c_p \geq c_{p,threshold} \\ 0 & : \quad c_p > c \wedge c_p < c_{p,threshold} \end{cases} \quad (10)$$

with $c_{p,threshold}$, a threshold value (specified empirically) that limits the minimum acceptable value c_p. If no value is sufficient, the output is set to 0, and the 3D model disappears. This is necessary because a user may intend to remove a marker and the object from the working surface. Thus, the 3D model must disappear.

IV. INTERACTION ADVANTAGES

In this section the advantages for interaction gained by the hybrid tracking solution are explained. Two tests were conducted to demonstrate these advantages. The section starts with an introduction of the utilized hardware and software. Next, the first test is presented that shows the advantages. The second test results in a set of quantitative measurements, which underpin the results of the first test.

The application example is an AR assembly support application, which presents an assembly sequence as a set of 3D models superimposed on the parts to be assembled. The user handles the parts to be tracked and regularly covers the markers.

A. Application setup

The application runs on a Windows 7 PC with Intel Xeon 3.6 GHz Processor and 6GB RAM. The graphics card uses a NVIDIA Quadro 5500 processor. Two Creative LiveCam Chat HD video cameras are implemented for tracking with a resolution of 1024 x 768 pixels at 30 fps. The cameras were arranged on the left and right position of the working area. The average distance between working area and both video cameras was 80cm.

Figure 7 shows an image of the application main window. WorldViz Vizard 4.02 (http://www.worldviz.com) was used as the development platform with the ARToolworks ARToolkit multi marker tracking for marker tracking (http://www.artoolworks.com/).

Vizard is a platform that facilitates the development of virtual reality applications. It uses Python as its programming language. ARToolworks is a plug-in for

Vizard. It provides functions for AR applications like image capturing, tracking, and spatially correct rendering. The optical flow tracking has been implemented with OpenCV 2.3 (http://opencv.willowgarage.com/wiki/), an open source computer vision library. It is written in C++ and provides functions to realize optical flow tracking. The optical flow tracking algorithm described above was developed as plug-in for Vizard in order to integrate this functionality with the AR application.

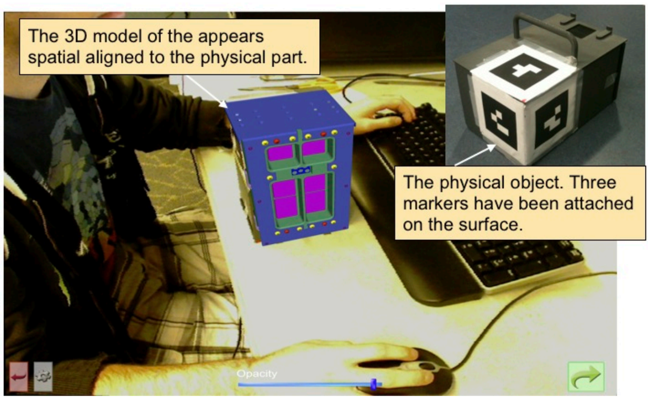


Figure 7. Main view of the AR application

B. Interaction test

The objective of this test was to assess the performance of the hybrid tracking in comparison to the common ARToolkit tracking. In the following, the term single-camera tracking refers to common ARToolkit tracking, i.e., the tracking of all markers using one camera only.

The question addressed by this test is: does the hybrid tracking facilitate manual interaction with an object to track? Hybrid tracking is assumed to be more robust than single-camera tracking: the physical object should be tracked continuously even when the markers are out of view of one camera or partially covered.

The physical object shown in Figure 7 (upper right corner) was used as a test object. Three ARToolkit markers were attached on the body of this object. Each marker has a size of 50mm in square. The AR application superimposes a video image with a 3D model of the assembly of that object.

Figure 8 shows images of the video sequences of the test. The images show the virtual object superimposed by a 3D model. The user rotates the objects in order to inspect it. A 90-second video sequence was used for this test. The video ensures equal input data for hybrid tracking and single-camera tracking.

During the test a user interacts with this object. The user grasps it with two hands, turns it, and inspects it from different orientations. Thus, the object changes its orientation to the camera continuously. While doing so the three markers are partially covered by the hands of the user.

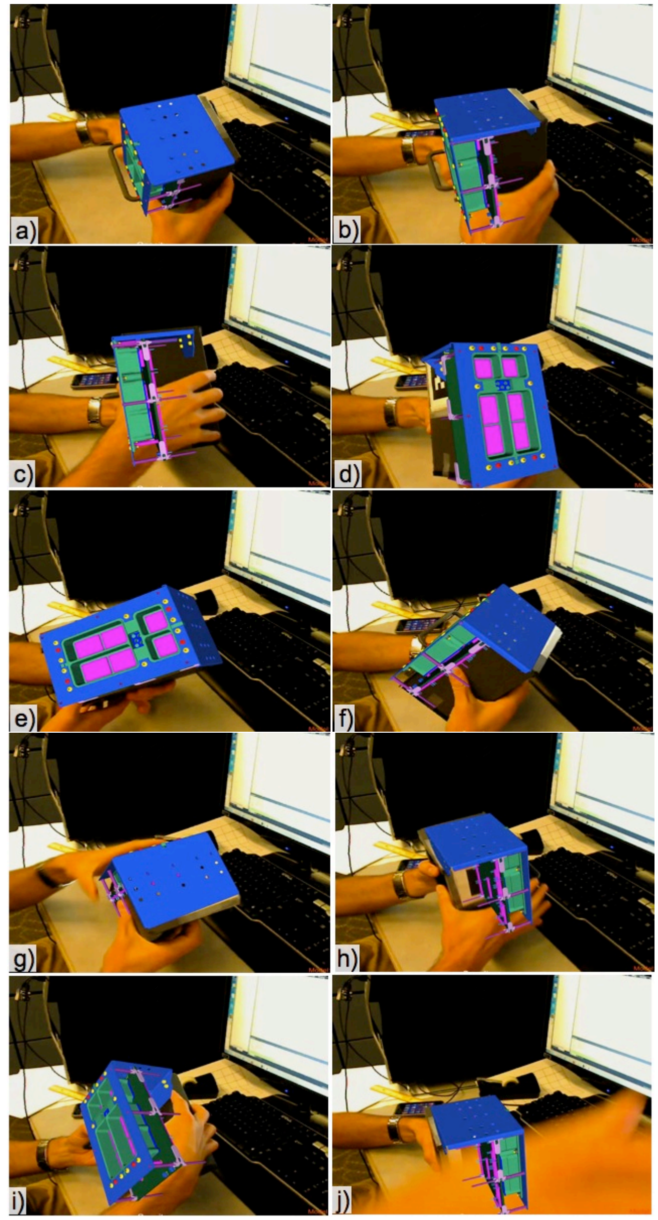


Figure 8. Image sequence retrieved from the test video

In Figure 8h) marker visibility to one video camera is partially covered by one hand of the user, however the 3D model was always visible and in its expected position.

To measure the robustness of the hybrid tracking and the single-camera tracking the confidence value c_n (Eq. 3) was recorded. Figure 9 shows a diagram with the results. The abscissa displays the number of frames, the ordinate the confidence value c_n. The upper graph depicts the results of the hybrid tracking, the lower graph the results of the single-camera tracking. The horizontal sectioning of the charts shows the threshold values, thus, it shows the mode of operations according to Eq. 10.

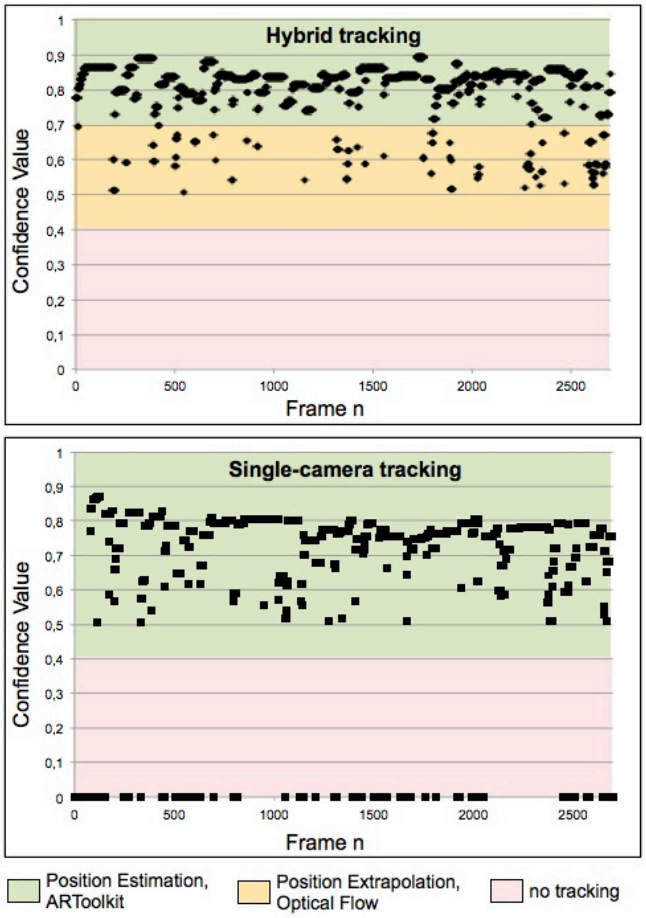


Figure 9. Confidence value of the a hybrid tracking test and a single-camera test.

For the hybrid tracking: the AR application switches between a position estimation mode (upper area), the position extrapolation mode (middle area), and an off mode when the confidence value declines beneath the given threshold. For the tests, a threshold of $c_p = 0.7$ and $c_{p,threshold} = 0.4$ were used. Below a value of 0.4 the AR application hides the 3D model. For the single-camera tracking: the threshold also was $\boldsymbol{c_p} = 0.4$.

The results show, the hybrid tracking can continuously track the physical object and show the 3D model. The confidence value moves within a range of 0.5 and 0.9. Thus, the AR application switches between the position estimation mode and the position extrapolation mode multiple times, but it never declines below the threshold value $c_{p,threshold}$. Thus, the 3D model never disappears during this test.

In comparison, the single-camera tracking is not able to track the physical object continuously. The graph shows that the confidence value occasionally declines below the threshold value c_p. In this case the AR application hides the 3D model.

C. Optical Flow Test

The objective of this test was to assess the performance and the advantages of the optical flow tracking. The optical flow tracking should be able to track the marker even if the marker is partially covered and the ARToolkit is not able to detect the 3D model. For this test a single ARToolkit marker was used that showed a semi-transparent 3D cube (Figure 10). The edges of the cube are aligned to the edges of the 3D model. Thus, an incorrectly calculated position and orientation of the cube can be observed.

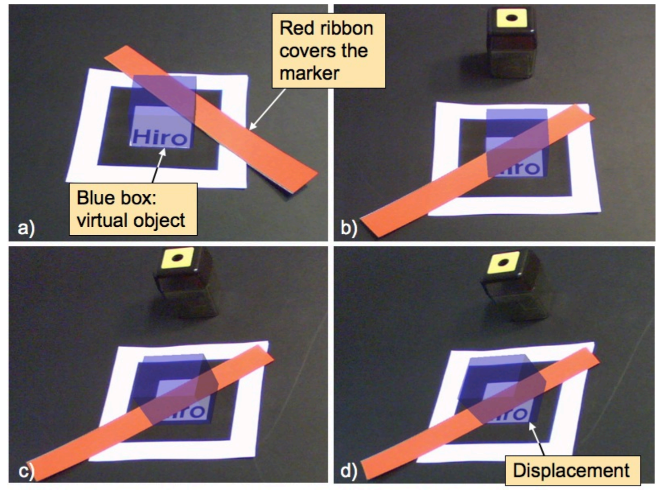


Figure 10. The optical flow tracking backs-up the ARToolkit marker tracking. Thus, partially covered markers could be tracked.

Initially, the position and orientation of the marker was calculated by the ARToolkit tracking. It was disabled after an initial transformation matrix was calculated. During the test a user covers the ARToolkit marker partially using a red ribbon. When doing this, the 3D cube must remain at its position (Figure 10). In addition, the camera was moved in one test (Figure 10 b-d). The images indicate that, in general, the tracking is able to estimate the position and orientation of the marker and to keep the 3D model at the expected position.

D. Discussion

The results of both tests are a strong indication of the interaction advantages gained by the hybrid tracking solution: it facilitates more natural interaction. The objective of an AR application is to allow natural interaction meaning, the user does not need to consider the technical solution of an interaction technology. In marker-based AR applications, interaction relies on the quality of tracking. Thus, for natural interaction, the user should not be asked to consider markers during interacting. The common marker-based solution does not allow that. The results presented in Figure 9, Single-camera tracking, indicates that the 3D model disappears when the user handles the physical object; the confidence value, which is the parameter responsible for switching between extrapolation and estimation, declines below the threshold value $c_{p,threshold}$. In general, it is also well known that the user must keep the marker in the line of sight of the video camera in order to keep the virtual information visible. In contrast, the hybrid tracking makes the interaction more natural. The 3D model does not disappear regardless of the interaction of the user (Figure 9, Hybrid tracking); the

 ISBN: 978-1-61208-250-9

confidence value never drops below the critical threshold value. Thus, the user does not need to consider the optical markers while working with a physical component.

The results of the second test show that it is possible to estimate the correct position for at least one second. Thus, the optical flow tracking is able to backup the ARToolkit when pattern recognition is not possible.

However, some limitations must be pointed out. First, the tests do not cover all possible situations and settings. Only one video with one parameter set has been tested. While this demonstrates the feasibility of the hybrid tracking, the reliability has not been demonstrated. However, experience with the ARToolkit alone in the demonstrated use case, indicates that hybrid tracking is a suitable solution for virtual assembly training.

Secondly, the optical flow tracking is a backup and works for a few frames only. It cannot be used as a second fully functional tracking system that replaces the ARToolkit. In addition, the quality of the tracking results decrease with time. This results in incorrect alignment and position of the 3D model. The entire approach works only when the movement of the marker between each frame is small. The optical flow tracking provides a linear approximation of the movement in a spatial coordinate system. The simplicity of the solution is an advantage at this point.

Thirdly, only three markers of one size were used to track a physical object. Thus, the results of hybrid tracking are limited to this setup. More markers can cause more disturbances because the software has to merge more data.

V. RESUMEE AND OUTLOOK

In this paper a hybrid tracking solution is presented that facilitates natural interaction in marker-based AR applications. Usually, interaction in marker-based AR applications relies on the handling of physical markers. Therefore, an unobstructed view between a video camera and the marker is mandatory. Tracking, and thus interaction, is impossible, if a user, his/her hands, or anything else covers the maker. In virtual training applications this regularly happens, because a user interacts with the physical objects that are subject to tracking. The hybrid tracking system developed uses two video cameras as input devices and combines the determined tracking data to one position and orientation estimation. In addition, an optical flow tracking approach is used to estimate the movement of a physical object when an ARToolkit marker is partially covered.

Thus, this paper makes two contributions: First, is a tracking solution that facilitates tracking of objects with which the user interacts. Hybrid tracking facilitates natural interaction with marker-based AR applicaitons.

Secondly, the optical flow tracking and a position estimation technique presented can act as a tracking backup for several seconds. The approach is simple from a technical point of view, compared with other tracking solutions, and tests indicate that it is feasible and necessary for virtual training applications.

Future work will address displacement problems and subassembly tracking. As mentioned above, the optical flow tracking and the switch between the position estimation and position extrapolation mode causes a displacement of the 3D model. Since the optical flow tracking becomes inaccurate with the time, this cannot be avoided. The goal is to replace the switch by a soft fade. Thus, the visual effect of the displacement will be reduced and it will be less significant for the user.

Assembly training also necessitates tracking of subassemblies. At this time only one physical part is tracked. To track subassemblies, ARToolkit markers may also be attached to each subassembly. To do this, several markers may be hidden. The hybrid tracking system must be enhanced to not consider this as an error.

VI. REFERENCES

[1] Azuma, R., 1997, “A Survey of Augmented Reality,” in Presence: Teleoperators and Virtual Environments Vol. 6 (4) (August 1997), pp. 355-385

[2] Blake J., 2011, “The natural user interface revolution,” Natural User Interfaces in .NET, pp. 1–33.

[3] Kato, H. and Billinghurst, M., 1999, “Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System,” in 2nd International Workshop on Augmented Reality, October 20-21, 1999, San Francisco, USA, p. 10

[4] Welch G. and Foxlin E., 2002, “Motion tracking: No silver bullet, but a respectable arsenal,” IEEE Computer Graphics and Applications, Nov./Dec. 2002, Vol. 22 (6), pp. 24-38

[5] Fiala, M., 2005, „ARTag, a fiducial marker system using digital techniques,“ in: Conference on Computer Vision and Pattern Recognition, 2005 (CVPR 2005), 20-25 June 2005, San Diego, CA, USA, IEEE Computer Society, Vol. 2, pp. 590 - 596

[6] Wagner, D., Langlotz, T., and Schmalstieg, D., 2008, “Robust and unobtrusive marker tracking on mobile phones.” In Proc. of the Int. Symposium on Mixed and Augmented Reality (ISMAR 2008), 15-18th September 2008, Cambridge, UK, pp. 121–12.

[7] Wagner, D. and Schmalstieg , D., 2007, "ARToolKitPlus for Pose Tracking on Mobile Devices", In Proceedings of 12th Computer Vision Winter Workshop (CVWW'07), 6-8 February 2007, St. Lambrecht, Austria

[8] Naimark, L. and Foxlin, E., 2002, “Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker,” In: ACM Int. Symposium on Mixed and Augmented Reality (ISMAR 2002), 30 September - 1 October 2002, Darmstadt, Germany. IEEE Computer Society 200, pp. 27–36

[9] Uchiyama, H. and Saito, H., 2011, “Random dot markers,” IEEE Virtual Reality (VR ‘11), 19-23 Mar. 2011, Singapore, pp. 35-38

[10] Seo, J., Shim, J., Choi, J.H., Park, C., and Han, T., 2011, “Enhancing Marker-based AR Technology,” in: Shumaker, R. (Ed.): Virtual and Mixed Reality, Part I, Proceedings of Human Computer Interaction International 2011 (HCII’11), 9-14 July 2011, Orlando, FL, USA, pp. 97-104

[11] Marimon, D., Maret, Y., Abdeljaoued, Y., and Ebrahimi, T., 2007, "Particle filter-based camera tracker fusing marker- and feature point-based cues," in: Proc. of the Conf. on Visual Communications and Image Processing, IS&T/SPIE Electronic Imaging

[12] Piekarski, W., Avery, B., Thomas, B., and Malbezin, P., 2003, "Hybrid Indoor and Outdoor Tracking for Mobile 3D Mixed Reality," in: International Symposium on Mixed and Augmented Reality (ISMAR2003), Oct. 7-10, 2003, Tokyo, Japan

[13] Kalkusch, M., Lidy, T., Knapp, M., Reitmayr, G., Kaufmann, H., and Schmalstieg, D., 2002, “Structured Visual Markers for Indoor Pathfinding.” In: 1st Int'l Augmented Reality Toolkit Workshop, 29. Sep. 2002, Darmstadt, Germany

[14] Yang, P., Wu, W., and Moniri, M., 2007, "A Hybrid Marker-based Camera Tracking Approach in Augmented Reality" In: Proceedings of the 2007 IEEE International Conference on Networking, Sensing and Control, 15-17 April 2007, London, UK, pp. 622-627

[15] Yang, H., Cho, K., Soh, J., Jung, J., and Lee, J., 2008, "Hybrid Visual Tracking for Augmented Books," in: Stevens, S.M. and Saldamarco, S. (Eds.): Proceedings of the 7th International Conference Entertainment Computing, ICEC 2008, Pittsburgh, PA, USA, September 25-27, 2008, LNCS 5309, pp. 161–166

[16] Fischer, J., Eichler, M., Bartz, D., and Straßer, W., 2006, "Model-based Hybrid Tracking for Medical Augmented Reality" In: Proceedings of Eurographics Symposium on Virtual Environments (EGVE), Lisbon, Portugal, pp 71-80

[17] Pensyl, W.R., Tran, C.T., Qui, Pei Fang, Hsin, Shang Ping Lee and Jernigan, D. K. , 2009, "Large Area Robust Hybrid Tracking with Life-size Avatar in Mixed Reality Environment – for Cultural and Historical Installation", The International Journal of Virtual Reality, 2009, Vol. 8, Issue 2, pp. 13-18

[18] Newman, J., Wagner, M., Bauer, M., MacWilliams, A., Pintaric, T., Beyer, D., Pustka, D., Strasser, F., Schmalstieg, D., and Klinker, G., 2004, "Ubiquitous Tracking for Augmented Reality", in: Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '04, Nov. 2 - 5, 2004, Arlington, VA, USA, pp. 192 - 201

[19] Su, H., Kang, B., and Tang, X., 2008, "Research and Implementation of Hybrid Tracking Techniques in Augmented Museum Tour System," In: Pan, Z. et al. (Eds.): Proceedings of the Third International Conference on Technologies for E-Learning and Digital Entertainment, Edutainment 2008, June 25-27, Nanjing, P. R. China, pp. 636–643

[20] Szeliski, R., 2010, "Computer Vision". Springer Verlag, Berlin

[21] Bradski, G. and Kaehler, A., 2008, "Learning OpenCV," First Edition, O'Reilly Media Inc., Sebastopol, CA

[22] Harris, C. and Stephens, M., 1998, "A Combined Corner and Edge Detector," in Fourth Alvey Vision Conference, Manchester, UK, pp. 147-151

[23] Welch, G and Bishop, W., 2006, "An Introduction to Kalman Filter," Technical Report TR 95-041, Department of Computer Science University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[24] Tsai, R., 1987, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," in: IEEE Journal of Robotics and Automation, Vol. No. 4, August 1987, pp. 323-344

Evaluating Multi-Modal Eye Gaze Interaction for Moving Object Selection

Jutta Hild*, Elke Müller*, Edmund Klaus*, Elisabeth Peinsipp-Byma*, Jürgen Beyerer*†
*Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB)
Karlsruhe, Germany
†Vision and Fusion Laboratory, Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
e-mail: {jutta.hild; elke.mueller; edmund.klaus; elisabeth.peinsipp-byma; juergen.beyerer}@iosb.fraunhofer.de

***Abstract*—Moving object selection is a frequently occurring interaction task in expert video analysis. State-of-the-art, video analysts use the mouse as input device. As the selection of moving objects is more complex than the selection of static objects, particularly when objects move fast and unpredictable, using the mouse is less efficient than usual and induces more manual stress. In this contribution, multi-modal gaze-based interaction is proposed as an alternative interaction technique for moving object selection. In an experiment using an abstract moving circle scenario, the two gaze-based interaction techniques gaze + key press and liberal MAGIC pointing are compared to mouse interaction. Evaluation of both user performance and user satisfaction shows that at least the gaze + key press technique might be a promising interaction alternative for moving objects selection.**

Keywords-moving object selection; eye gaze interaction; multi-modal interaction; experiment; video analysis

I. INTRODUCTION

The selection operation is one of the basic interaction tasks in human-computer interaction for desktop computer systems with graphical user interfaces (GUI). To perform a selection task, most users use the mouse as an input device because it provides an intuitive and effective interaction technique for the selection of typical GUI elements of different sizes as, e.g., windows, icons, or buttons. However, there are applications that also require the selection of moving objects. E.g., in live-video analysis for security applications moving object selection is a frequently occurring interaction task.

A. Moving object selection in video analysis

Basically, video analysis is a visual search task aiming to detect and analyze objects and situations for conspicuous events, and to report them to the authorities concerned with the response actions. Therefore, video analysts are equipped with video exploitation systems providing a large representation of the scene currently by the video sensor. Figure 1 shows, as an example, the ABUL user interface [1]. Typically, analysts use the visualization in the large window in the center to perform their search task by continuously scanning it for conspicuous events. When detecting an event, the analysts have to mark it in the video stream and send the labeled video clip to the concerned authorities for further investigation/examination.

But, due to motion of the video sensor all objects including static ones like buildings move, too. Hence, object selection using mouse interaction can be rather challenging. The reason for this can be revealed easily considering the selection process for pointer-based input devices like mouse, consisting of the three parts:

1. Picking the object to be selected (by gaze)
2. Pointing at the object by moving the selection device onto the object (by motor action to position the mouse pointer)
3. Actuating the selection (by motor action to perform a mouse button click)

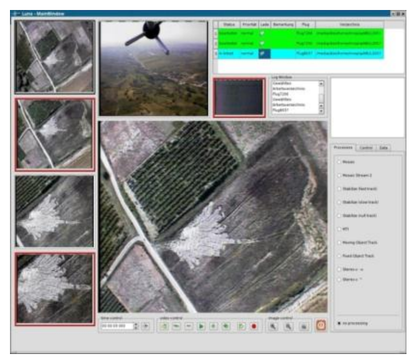


Figure 1. ABUL user interface tailored for moving object selection.

If a static object has to be selected the three parts are each performed once. If a moving object has to be selected it might be necessary to perform at least part two more than once, as the object might move away from the mouse pointer position before the user is able to perform the mouse click. There may be even a number of pick-point-click-loops necessary before the user finally is able to accomplish the moving object selection. This can happen, e.g., if due to high wind the video sensor, and accordingly the video image is shaking. Additionally, as well due to video sensor motion, objects are visible on the screen only for a limited time span, and accordingly selectable only for a short time. Thus, the analyst has to perform the moving object selection very fast, especially, if an object moves in addition to the sensor motion, as for example driving vehicles do.

 ISBN: 978-1-61208-250-9

However, as a missed object selection in security applications might result in severe consequences, it is important that the analyst is able to reliably select the moving objects. As a consequence, the load for the hand and arm motor systems is rather high. According to statements made by expert video analysts, their working time often lasts for several hours. Hence, mouse interaction often induces repetitive strain injury. For these reasons it is worth developing an alternative interaction concept which allows moving object selection, on the one hand, in a more efficient way, and, on the other hand, with reduced manual stress.

B. Proposing eye gaze interaction for moving object selection

The alternative which is proposed in this contribution is eye gaze interaction. Technically, eye gaze interaction is facilitated by an eye tracking device which provides continuously the current gaze position of the user on the screen. Over the years, a lot of different eye gaze interaction variants have been proposed [2,3,4,5]. They can be subdivided into unimodal and multimodal interaction techniques. While with the unimodal techniques the whole selection process described above is performed using eye gaze input, multi-modal eye gaze interaction techniques use at least one other input modality.

However, both unimodal and multimodal techniques have in common that they use eye gaze position for pointing. As a result, parts 1 and 2 of the mouse selection process described above are performed together, at the same time. This property is well suitable for moving object selection in video analysis. As introduced before, video analysis is basically a visual search task and displayed objects might be visible and selectable only for a short time. Therefore it is essential to ensure that the analysts can keep their visual focus of attention continuously on the object during the selection process. This is possible using eye gaze interaction. But it is not for mouse interaction where the visual attention has to be focused onto the mouse pointer when looking for it and repositioning it. Besides the advantage of not having to divide the visual focus of attention between object and pointing device, to perform parts 1 and 2 together should result in an overall shorter selection time. This could be beneficial if an object is visible in the video only for a short time or if its position changes quickly and unpredictably. Furthermore, as the user has no longer to move the pointing device by hand-motor action the user's manual stress is reduced, which could contribute to a decrease of hand motor complaints.

What makes the eye gaze interaction techniques different is the way part 3, the equivalent to the mouse click equivalent, is performed. Examples of uni-modal eye gaze interaction techniques use, e. g., a certain eye gaze dwell time on the object, or an eye blink [3]. Examples of multi-modal eye gaze interaction techniques use, e.g., a button press on a keyboard [2,6,7], manual pointing [4], or even speech [8]. All of them have their advantages, but as our goal is to provide an interaction concept for a, both mentally and physically, stressful working task, techniques that would add additional load are not appropriate.

For this reason, both dwell time and eye blink are not appropriate as both require unnatural eye movement. Furthermore, as stated, e.g., in [4], one should be careful to load gaze with a motor control task as it has been evolved for visual perception. Hence, performing a motor control task by eye gaze would result in additional mental load. Besides, to overload the eye normally used for information input with an information output task results in the so called "Midas Touch" Effect. As the eyes are an always-on device it is necessary to ensure by appropriate interaction design to make eye movements for perception distinguishable from eye movements for pointing or selection [3,5]. Therefore, for our concept eye gaze shall not be used for the final selection actuation.

As the multi-modal interaction techniques use an input modality for selection actuation different from eye gaze they do all avoid the Midas Touch effect [7]. Furthermore, both hand and speech are natural means for object selection. As state-of-the-art video exploitation systems not necessarily comprise a speech recognition system we did not consider speech as input modality at this time. The button press on a keyboard being referred to as "hardware button" [2], "Gaze and Keyboard" [5], "Eye tracking with manual click" [7], or "Eye + Spacebar" [6], has been reported by these (and other) authors to be a rather fast, easy-to-use, and effortless interaction technique. In addition to this, several times it was reported to be the favorite technique of subjects for selection tasks in experiments [5,6]. Because of these good assessments for static object selection, this technique was selected to be one of the appropriate candidates for performing moving object selection and was therefore evaluated in our experiment. From now on, we will refer to it as the "gaze + key press" technique (GK).

However, there is one major drawback of techniques using eye gaze for pointing. On the one hand, due to eye tracking inaccuracy and, on the other hand, due to anatomic and physiological properties of the eye, it is not possible to determine the exact gaze position on the screen. Eye tracker manufacturers often specify an accuracy of 0.5°. This means, whatever gaze position is estimated by the eye tracker, the true position might be located within a circle of 1° of visual angle around the estimated position. Furthermore, as the region of sharp vision, the fovea, covers 1° to 2° of visual angle, and as it jitters continuously to keep the visual stimulus on the retina alive it represents a "region of uncertainty" for the eye tracker [9]. To get higher accuracy, in [4] a gaze-based interaction technique called MAGI pointing, combining eye gaze and manual interaction, was proposed. The concept is to use eye gaze for coarse pointing, and a pointer-based interaction technique performed by hand for fine pointing and selection actuation. As one of the MAGIC pointing approaches, the so called liberal MAGIC pointing, performed faster than mouse interaction in a static object selection task, it was also considered to be an appropriate candidate for performing moving object selection and was therefore evaluated in our experiment. From now on we will refer to it as the "MAGIC liberal" technique (MAGIC-lib).

II. RELATED WORK

The challenge of moving object selection using mouse interaction has been addressed before by several authors. A comprehensive survey on the issue is provided by [10], including considerations on moving object selection in videos. Discussed techniques for adaption and improvement of mouse interaction can be categorized as techniques that extend the activation area of the mouse pointer according to object size [11] and techniques that extend the activation area of the object itself [10], both being reported to provide improved moving object selection using mouse.

Selection with eye gaze has been widely examined for static objects. As stated in the introduction, the GK technique was part of many evaluations [2,5,6,7] and found to be faster than mouse, intuitive, easy-to-use, and manually little stressful. However, selection accuracy was much lower than mouse, resulting in higher selection error rates. The MAGIC pointing paradigm was also evaluated in various variants by several authors and reported to be a good mouse alternative [4,12].

Contributions comprising gaze-based moving object selection are available for computer gaming applications. In [13] a survey is presented proving taxonomy of game genres and eye gaze interaction applicability. Two genres, shoot-them-up games and first-person shooter games, comprised moving object selection (shooting). When gaze was used as an input for first-person shooters (gaze being used for crosshair control, mouse click for triggering the weapon fire) studies showed that gaze input could not beat mouse input [13,14]. Targeting the objects with gaze provided much lower accuracy than the mouse. Another study used gaze as an input device for a chicken shoot game [13]. After a few practice trials, the participants were able to perform much faster with eye controlled shooting (combination of controlling the cross-hair by gaze position and triggering weapon fire by mouse click) than with mouse and keyboard.

III. EXPERIMENT

To examine whether, for moving object selection, the gaze-based interaction techniques gaze + key press (GK) or liberal MAGIC pointing (MAGIC-lib) would provide a promising alternative to mouse interaction an experiment was designed. To compare the three interaction techniques, a simple moving object selection task was designed simulating the real moving object selection in video analysis as described in the introduction. For every interaction technique, performance was measured by determining selection completion times and selection error rates. Satisfaction was evaluated using a questionnaire asking to rate selection accuracy, selection speed, task adequacy, and user friendliness. In addition, participants were asked for their favorite interaction technique for moving object selection.

A. Interaction Techniques

Object selection by mouse interaction used the typical point-and-click-procedure with a left mouse button click. With the GK interaction technique, object selection required to look at the object and to actuate the selection by pressing the ENTER-key while fixating the object. MAGIC pointing was implemented using the liberal approach. The pointer was continuously visualized at the currently measured fixation position using the I-DT fixation algorithm [15]. For selection, the mouse had to be moved over the object, and the left mouse button had to be clicked, just in the standard mouse interaction. The change of pointer control between eye tracker (gaze) and mouse (hand) worked as follows. By default, the control of the pointer lay with the eye tracker. In case of mouse movement, the pointer control was transferred immediately to the mouse. Pointer control was transferred back to the eye tracker only after the mouse had not been moved for 100 ms.

B. Apparatus

To capture the gaze data, a table-mounted Tobii 1750 remote eye tracker was used. It is a video-based eye tracker using pupil-center and corneal-reflection for gaze measurement. The eye-tracking hardware (video cameras and near infrared LEDs) is integrated with the display frame of a 17-inch TFT display with a 1280x1024 resolution. According to manufacturer information, the Tobii 1750 features an accuracy of 0.5°, and a spatial resolution of 0.25°. Gaze position is sampled at 50 Hz.

Participants sat at a distance of 60 cm from the monitor which according to the manufacturer allows a freedom of head-movement of 30 x 16 x 20 cm (W x H x D). Therefore, no chin-rest was used in order to let participants use the eye tracker with the same freedom of movement that would be required for video analysts at work.

Mouse input for mouse interaction and MAGIC-lib interaction was performed using a standard optical mouse. To perform the key press for the GK interaction, a standard keyboard's ENTER-key was used.

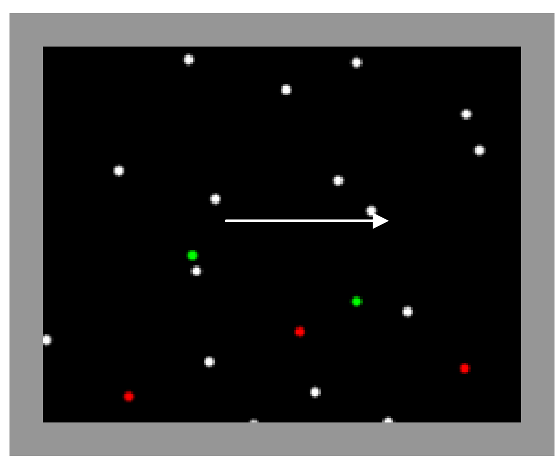


Figure 2. Screenshot of the experimental task. The arrow shows the moving direction.

C. *Experimental Task*

The moving object selection task was designed referring to movements of persons or vehicles in real world surveillance videos. Figure 2 shows a screen shot of the task. The objects were represented by white circles. All had the same size of a diameter of 24 pixels (corresponding to 0.63 cm on the used monitor, and 0.66° of visual angle). The objects moved over the screen from left to right on a black background. Each object moved with one of three constant speeds resulting in a visibility on the screen of about 7 to 9 seconds. While moving over the screen, some prior to the start of the experiment randomly chosen objects were highlighted in red for one second. In total, 294 objects moved over the screen, 91 of them were highlighted. To induce some of the time pressure and stress video analysts are faced with, the number of objects on the screen steadily grew during the task. In total, to perform the test task once lasted for about 180 seconds. The task was to select the highlighted objects. However, all objects were selectable. After being selected, an object changed its color into green as a visual feedback for the user.

As pointing based on gaze position can provide only limited accuracy, particularly, when allowing some head movement, a radius of acceptance of 80 pixels was assigned to the objects. Thus, the visible size of the circles was 24 pixels, while their selectable size had a diameter of 160 pixels (3.97° of visual angle). This approach is related to the object magnifications proposed, e. g, by [10] to improve mouse interaction. However, in contrast to their suggestions, our approach uses an invisible object magnification. This is more suitable for video analysis applications as they usually do not allow any part of the video to be covered by artificial visualizations. Regarding real applications, our radius of acceptance approach would require information about the position coordinates of the objects. This requirement is met as state-of-the-art video exploitation systems like ABUL [1] feature automatic target recognition algorithms which are able to preprocess the video material and to provide exploitation results like object position coordinates.

D. *Participants*

20 participants volunteered for this study. All of them were university students or members of our department and were therefore experienced users of desktop computer and mouse. Two of them could not complete the tasks due to calibration problems and were therefore excluded from the data analysis. The age of the remaining 18 participants (16 male, 2 female) ranged from 18 to 54 years (mean = 28.6 years). All participants had normal or corrected to normal vision, two of them wore glasses, three used contact lenses. Nine participants had already used the Tobii 1750 eye tracker once before.

E. *Procedure*

For the study, a within-subjects design was used. Each participant performed the selection task three times, each time using a different interaction technique. To control fatigue and training effects the participants were divided into 6 groups, each group performing the experiment with a different order of techniques, thus following a complete, counterbalanced design.

The experiment started with an explanation of the moving object selection task using a short practice task of about 50 moving circles. Participants were told to perform selections as fast as possible and with the least possible number of mistakes. In case of performing one of the two gaze-based interaction techniques, the participants completed the standard Tobii 1750 calibration using a grid of 9 calibration points before practicing. Participants were allowed to repeat the practice task until they felt to be familiar with the interaction technique. After that, in case of a gaze-based interaction technique a recalibration had to be completed, again using the 9-point-calibaration-procedure. Then, the test task of 294 circles was performed. Finally, the participants rated rate selection accuracy, selection speed, task adequacy, and user friendliness on a 5-point scale (1: best rating, 5: worst rating) and reported their favorite interaction technique for moving object selection.

IV. RESULTS

Figure 3 shows the results for the selection completion time (SCT). SCT was measured by the difference between the time of selection and the time of highlighting. A repeated-measures analysis of variance (ANOVA) shows that the participants' performance significantly varied with techniques ($F(2;50) = 510.7$; $p<0.001$). A post-hoc analysis with Bonferroni correction shows high significant differences in SCT between GK and both of the other techniques ($p<0.001$), and reveals a significant result between the SCT of mouse and MAGIC-lib ($p<0.01$). With an average SCT of 786 ms the GK technique was considerably faster as the mouse interaction technique with an average SCT of 1422 ms. Participants performed slowest with MAGIC-lib (average SCT = 1681 ms).

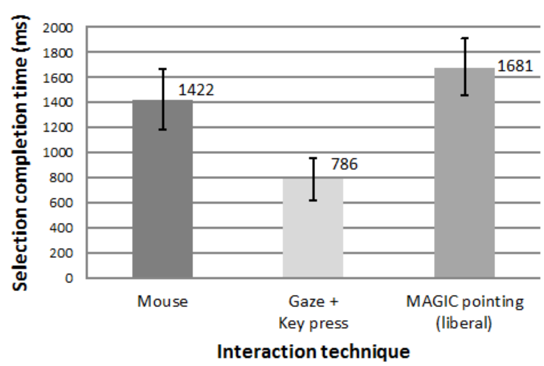


Figure 3: Selection completion time as a function of interaction technique.

Figure 4 shows the results for the selection error rate ((number of wrong selections / 294) * 100). It varies significantly with techniques as well ($F(2;50) = 17.41$; $p< 0.001$). Bonferroni-corrected comparisons show that the error rate for the MAGIC-lib technique is significantly higher than for GK and mouse ($p<0.0001$). The difference between GK and mouse is not significant. Participants performed with an average error rate of 3.3 % using mouse, with an average error rate of 3.9 % using GK, and with an average error rate of 8.3 % using MAGIC-lib. Error rates were similar for mouse and GK, whereas they were about double for MAGIC-lib.

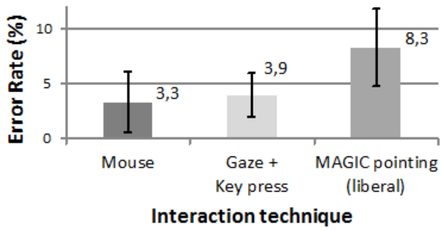


Figure 4. Selection error rate as a function of interaction technique.

Analyzing the errors in more detail, Figure 5 shows the percentage of missed objects (highlighted objects that could not be selected) and the percentage of false alarms (non-highlighted objects that were selected by mistake). An ANOVA of the number of missed objects shows a significant result (F(2;50) = 21.55; p<0.001). The Bonferroni post-hoc correction reveals significant differences between missed objects of MAGIC-lib and both GK and mouse (p<0.001). There is no significance between GK and mouse. In absolute numbers, participants missed with mouse on average 8.4 objects (out of 91), with GK 7.5 objects, and with MAGIC-lib 22.0 objects. Again, participants performed similar using mouse and GK and using MAGIC-lib considerably worse.

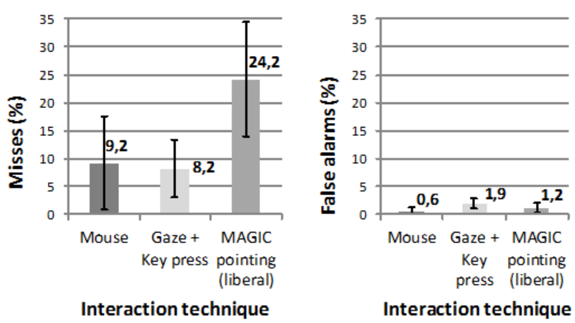


Figure 5: Misses and false alarms as a function of error rates.

An ANOVA of the number of false alarms shows a significant result as well (F(2;50) = 11.02; p < 0.001). Significant differences of false alarms between mouse and GK (p<0.001) are revealed by the post-hoc analysis with Bonferroni correction. At a p = 0.05 level a significant result is shown between mouse and MAGIC-lib. No significance is given between GK and MAGIC-lib (p=0.153). In absolute numbers, using mouse there were on average 1.3 (out of 203) objects selected by mistake, using GK it were 3.9 objects, and using MAGIC-lib it were 2.4 objects.

Figure 6 shows the results of the ratings of user satisfaction from the questionnaire. For all categories, MAGIC-lib got the worst ratings. For selection accuracy, the participants rated mouse and GK almost equally good. For selection speed, task adequacy, and user friendliness, GK was rated best, followed by mouse and MAGIC-lib.

Finally, asked for their favorite interaction technique for moving object selection, 16 out of the 18 participants voted for GK, the other two for mouse interaction.

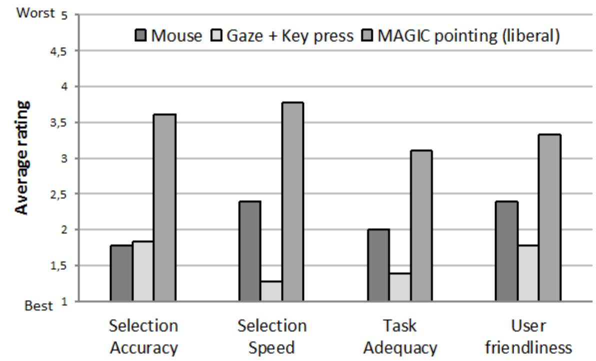


Figure 6. Subjective ratings for user satisfaction.

V. DISCUSSION

As shown for the selection of static objects [2,5,6,7], the GK technique also provides a very fast technique for the selection of moving objects. In the experiment presented here, participants were able to perform much faster than with mouse interaction. While the liberal MAGIC pointing approach provided a little faster object selection for static objects as mouse [4], this was not the case for moving object selection. Only one participant was able to select faster with MAGIC-lib than with the mouse. On the other hand, MAGIC pointing is a technique which is close to mouse interaction – usually at its end at least a small mouse interaction is completed. It is clear that it is rather challenging for the participants, both mentally and due to slightly different hand motor action, to use a technique that is almost like the one they are used to. Due to this fact, as already mentioned in the description of the procedure of the experiment, participants were allowed to repeat the practice task with each technique as often as they wanted. But, of course, even practicing for several minutes cannot compensate years of mouse experience. Moreover, some participants felt distracted by the permanent visibility of the mouse cursor. This could, of course, also have increased the selection completion time. The subjective estimations for selection speed in the questionnaire confirm the results of the performance measurement. All participants rated GK being the fastest selection technique. After all, 3 participants rated MAGIC pointing to be faster than mouse.

Considering the selection error rate, mouse and achieve an equally good error rate. Of course, this is due to the rather generous radius of acceptance which surrounded the circles. For static objects, it has been reported that for a size of 2 cm and larger [7], GK provides rather good selection accuracy. On the other hand, in video analysis it might in many cases be sufficient to select an object within a radius of uncertainty. Again, participants performed worst using MAGIC pointing. But again, this might be due to its complexity and its closeness to mouse interaction. After all, at least two participants were able to perform with the fewest number of selection errors using MAGIC pointing. As for selection speed, the ratings from the questionnaire for selection accuracy meet the results from the performance measurement. However, considering the error rate in more detail shows that there is at least one parameter that refers to the uncertainty of eye gaze pointing as use in the GK technique: the false alarm rate is, even if rather low for all the techniques, highest for GK. The result for the misses is in line with the results of the total error rate.

The ratings for task adequacy and user friendliness were also best for GK which make it the most comfortable technique to use. Regarding mouse interaction, some participants noted that even within the short time of the experiment the selection of moving objects was very fatiguing using mouse.

VI. CONCLUSION AND FUTURE WORK

The experiment showed that for moving object selection gaze-based interaction can provide advantages. Overall, the GK technique holds the largest number of positive characteristics. It is fast and, as a consequence, few objects are missed. It is easy-to-learn, requires only little manual effort – an important characteristic for video analysts working for several hours –, and easy-to-use which all in all made the participants to prefer it for moving object selection. Considering the work of a video analyst, particularly a short selection completion time is important. The less time the analyst needs to select an object the earlier they can draw their attention to the next object. However, the accuracy of GK was so well due to the acceptance radius around the circles. Hence, there is a need to confirm selection accuracy similar to mouse accuracy for moving object selection using real scenarios, including unpleasant selection situations like close objects, small objects, or shaking videos. In this case, using a large radius of acceptance could result in an overlap of selection regions of objects. This would prohibit to reliably select a certain object unambiguously. The performance of the standard input technique, the mouse interaction, was not as fast as GK interaction. But, at least mouse interaction was fast enough to catch almost as much highlighted objects as the very fast GK technique. However, only after the short experiment, participants noted that moving object selection using mouse interaction is very fatiguing. Regarding MAGIC pointing, the participants did not get along with this technique as good as with mouse or GK. However, there were persons performing fast and accurate using this technique. Bearing in mind the distraction of some participants due to the permanent display of the pointer using MAGIC pointing it could be interesting to investigate other variants like the conservative approach or MAGIC touch [12] for moving object selection.

ACKNOWLEDGMENTS

The underlying project to this article is funded by the WTD 81 of the German Federal Ministry of Defense under the promotional reference E/E810/BC013/AF144. The authors are responsible for the content of this article.

REFERENCES

[1] N. Heinze, M. Esswein, W. Krüger, G. Saur. “Image exploitation algorithms for reconnaissance and surveillance with UAV.” In Proceedings of SPIE 7668, Paper 76680U, Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VII: 7 April 2010, Orlando, Florida, United States. Bellingham, WA: SPIE, 2010.

[2] C. Ware and H. Mikaelian, “An Evaluation of an Eye Tracker as a Device for Computer Input”, ACM SIGCHI Bulletin, vol. 18(4), pp. 183-188, 1987.

[3] R. J. K. Jacob, “The use of eye movements in human-computer interaction techniques: what you look at is what you get.” ACM Transactions on Information Systems (TOIS) 9, vol. 2, pp. 152-169, 1991.

[4] S. Zhai, C. Morimoto and S. Ihde, “Manual and gaze input cascaded (MAGIC) pointing.” In Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, pp. 246-253. ACM, 1999.

[5] M. Kumar, A. Paepcke and T. Winograd, “Eyepoint: practical pointing and selection using gaze and keyboard.” In Conference on Human Factors in Computing Systems: Proceedings of the SIGCHI conference on Human factors in computing systems, vol. 28, pp. 421-430, 2007.

[6] X. Zhang, and I. MacKenzie. “Evaluating eye tracking with ISO 9241-part 9.” Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments, pp. 779-788, 2007.

[7] R. Vertegaal, “A Fitts Law comparison of eye tracking and manual input in the selection of visual targets.” In Proceedings of the 10th international conference on Multimodal interfaces, pp. 241-248. ACM, 2008.

[8] D. Miniotas, O. Spakov, I. Tugoy, and I. S. MacKenzie. “Speech-augmented eye gaze interaction with small closely spaced targets.” In Proceedings of the 2006 symposium on Eye tracking research & applications, pp. 67-72, ACM, 2006.

[9] I. S. MacKenzie, “An eye on input: research challenges in using the eye for computer input control.” In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, pp. 11-12. ACM, 2010.

[10] H. Khalad, T. Grossman and I. Pourang, “Comet and target ghost: techniques for selecting moving targets.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI, pp. 839-848, ACM, 2011

[11] T. Grossman and R. Balakishnan, “The bubble cursor: enhancing targe acquisition by dynamic resizing of the cursor’s activation area.” Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 281-290, ACM, 2005.

[12] H. Drewes and A. Schmidt, “The MAGIC touch: Combining MAGIC-pointing with a touch-sensitive mouse.” Human-Computer Interaction–INTERACT 2009, pp. 415-428, 2009.

[13] P. Isokoski, M. Joos, O. Spakov, and M. Benoît, “Gaze controlled games.” Universal Access in the Information Society 8, vol. 4, pp. 323-337, 2009.

[14] J. D. Smith and T. C. Graham. “Use of eye movements for video game control.” In Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology, ACM, 2006.

[15] D. D. Salvucci and J. H. Goldberg. “Identifying fixations and saccades in eye-tracking protocols.” In Proceedings of the 2000 symposium on Eye tracking research & applications, pp. 71-78, ACM, 2000.

Touch-Screens and Elderly users: A Perfect Match?

Alma Leora Culén
Department of Informatics
University of Oslo
P.B. 1080, 0316 Oslo, Norway
almira@ifi.uio.no

Tone Bratteteig
Department of Informatics
University of Oslo
P.B. 1080, 0316 Oslo, Norway
tone@ifi.uio.no

Abstract—**This paper discusses some challenges in use of touch–based screens by elderly adults. We are focusing primarily on touch-based interactions with personal artifacts such as smart phones and tablets or touch-screens embedded in the home environment. We have conducted several small studies as a prequel to a larger study of "smart home" package designed and employed in "care residences" for elderly. We report here on findings from these studies and extend them into more general discussion on the use of touch interfaces by elderly. We discuss challenges related to diversity of elderly as a user group, progressive changes due to aging and their effects on the use of touch-screens. How to use technology to support mastering of daily life tasks and at the same time easy to use, touch-based solutions that support mastery (usually requiring some level of skills)? How to select other modes of interaction when the touch is not enough? Elderly people constitute a challenging and vulnerable user group that we want to strengthen and empower in the spirit of participatory design.**

Keywords—Touch Interfaces; Design Method; Elderly Users; Multimodal Interactions.

I. Introduction

We are experiencing dramatic trends in the diffusion of information and communication technology (ICT). A persistent increase in the ageing population is seen as both a challenge and an opportunity for ICT. Elderly adults are a diverse group, diversity further augmented by the nature of ageing. Gradual decline in visual and auditory perceptions, motor skills and cognitive abilities make elderly into a highly non-homogeneous user group in terms of physical and cognitive abilities. This has implications for interface design [7, 15].

Changes associated with ageing make elderly adults, as an ICT user group, more challenging and vulnerable than younger adults. The term '*vulnerable*' indicates that they may experience difficulties in raising their voice or getting their voice heard. The term '*challenging*' refers to the difficulties for designers to empathize with their experience of the world [11] given their gradual decline in physical ability.

A. Coming of Age

An important aspect of ageing is the process of losing physical and cognitive abilities over time. The process starts early (in the 20ies), but it is not until too much of the ability is lost that we see it as a problem. Progressive changes, in many cases, lead to increased need for care and loss of independence in later years. Most elderly prefer to live independently in their own homes as long as possible and postpone moving to senior communities or care centers [10]. The main indicators for having to move to a care unit are fall accidents, cognitive decline and loneliness [12]. Maintaining everyday physical and cognitive activities are crucial for not losing them and for counteracting these three problem areas. Technologies can replace or enhance lost abilities e.g. by smart home technologies (safety alarms, fall sensors, stove alarms) and by easy-to-use communication devices (video calls (skype or similar), easy-to-use phones such as the one shown in Fig. 1).

European Union has a large research initiative on Ambient Assisted Living [1]. Many national research-funding agencies all over the world invest into similar projects concerned with various aspects of designing technological solutions for and with elderly that are intended to provide assistance in their daily lives, increase the quality of life and prolong their stay at home (see [2, 3, 5, 8, 10, 13, 20]).

Figure 1. The phone designed for elderly, DoroS1, popular with elderly.

B. Touch Interfaces

There is a common perception that touch interfaces are intuitive, natural and easy to use, even for elderly [23, 28]. Smart phones, tablets and large multitouch surfaces are now commonplace and affordable. Thus, they are often the interface of choice for assisted living research and design.

 ISBN: 978-1-61208-250-9

However, there is very little systematic research undertaken on if and how these interfaces may enable and help elderly in interacting with their technology rich home environments [2]. On the other hand, there is an increasing amount of literature on principles and frameworks for touch interaction in general. Hornecker and Buur [14] present a framework emphasizing the interweaving of the material/ physical and the social, contributing to understanding social user experiences of tangible interaction. Similarly, Challis [4] discusses tactile interaction and its general principles. Wang and Quek discuss [28] how touch screens enable (intimate) contact and communication, and encourage exploration.

C. Do Touch Screens Fit Elderly Users?

In a newly started project: *'technology@home'*, we aim at studying and designing for elderly people living in their own apartments in a "care residence" (offering some services like a 24/7 reception desk, a restaurant, a gym to the inhabitants). The care residence offers a basic 'smart home' package to its inhabitants and we will study how the technology solutions are used and how they can be improved or enhanced through participatory design (PD). One of the apartments is reserved as a 'living lab' where inhabitants, relatives, carers and health care workers can get information, try out and learn about available technologies. In addition, the 'living lab' is an arena for PD experiments for new technical solutions.

We have used touch interfaces in earlier research with "challenging" users [6]. The 'smart home' package we will work with includes a touch screen that can be hung on the wall or be taken down and used as a tablet pc. Since there already are a number of apps for health care on tablets, we see a potential in evaluating relevant apps as a part of the process of designing new apps.

We are committed to work, starting in January 2013, on development of novel interfaces and services utilizing touch screens and tangibles, with users' mastery as a key design principle [3]. Elderly users (and their relatives and carers) will be recruited for collaboration and engagement in the (PD) design activities and evaluation sessions aimed at novel products and services (see [16, 19, 20]).

In this paper, we report from a set of small studies concerning the use of touch interfaces by elderly. These studies were carried out in preparation for the larger study described above, *technology@home*. The studies aimed to identify challenges and opportunities of touch technology for elderly people, as a basis for designing usable systems for and with this vulnerable user group.

II. Our Studies with Elderly and Touch-screens

This section presents our short studies of use and design on touch –based platforms with and for elderly people. The first example is a small study testing touch screen-based phones on a small set of elderly users. The three remaining examples are all semester long student projects [18, 25, 30], illustrating important challenges: how to balance simplicity and mastery, and how to make a good choice of interface when touch alone is not enough and multimodal interfaces may be called for.

A. Smart Phone Experience

For this experimental study we contacted a small sample of elderly with the aim to discuss and, if possible, observe their use of touch interfaces. Surprisingly, none of our participants have used touch interfaces, not even on a smart phone. Most of the elderly were using phones like the one shown in Fig. 1, designed with large buttons and straightforward functions. Friends and family often recommend the phone to them. It is sold in specialized stores targeting people with special needs.

None of the elderly we talked with liked this phone. The phone is bulky and heavy. Elderly, just like youngsters, like certain kinds of technology. Their tastes may be more along side of practical, but they do retain the sense of what is cool as a piece of technology and what is certainly not. However, they were familiar with them and liked the simplicity of use.

In order for them to become familiar with the new technology, they received touch-based phones and got some instructions for using them (see Fig. 2).

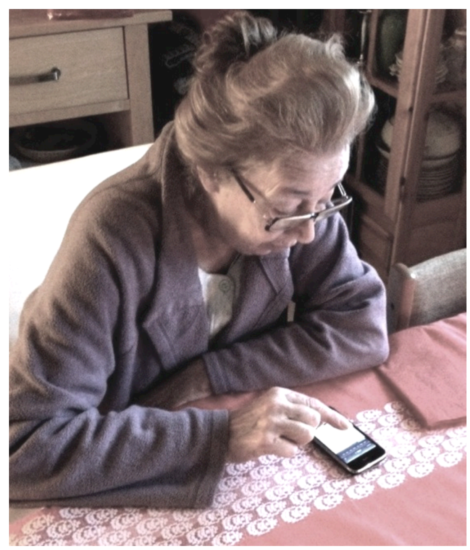


Figure 2. An elderly learning how to use a smart phone.

At the follow up meeting, the SIM card from one of the smart phones was back in the old phone. The elderly woman was very proud that she has managed to take it out of the new phone and place back into the old one. The explanation was that she does not need all the fancy stuff, she only uses the call functionality, does not feel like having Internet access, or any other features. When asked if the interface was too complicated on the new phone, the answer was that it was not difficult, it was easy and intuitive, but at the same time the smart phone was not needed. She was used to her old phone and attached to it emotionally, but had rational reasons for continued use. She argued that her old phone was smaller and lighter then the smart phone she got to try (android based one). The two functions that she uses, call and receive call, were very simple to use. She does not use the SMS feature. At first it was because there were three letters per small button, thus hard. She now does not wish to learn how to do it on the touch phone. If she did, SMS could become a new communication channel for her friends and family and she did not want to engage in that.

This example illustrates how important it is to listen to the users' reasons for using the technology or not (see also [17, 21, 23]). The use practices and habits are very important to consider when designing with elderly, and may be hard to break even when new solutions are easy, fun and intuitive. The design challenge is to provide space for users to keep the technology they use and like, but offering at the same time the possibility to master new devices and encourage an organic change of the established habits enabling them to master tasks and do things in easier ways.

Similar to the attachment to the old phone, many elderly still enjoy walking to the bank to pay their bills. Learning how to do it online may become a necessity at some point, as many banks already have only net-based solutions.

B. *TV control on the iPad*

Our next example is a student project [18] that studied how elderly people use their regular TV remote control, and aimed to design a touch-based solution based on knowledge about use practices. Four elderly people were involved in the project, but interesting challenges were encountered even within such a small sample.

The design team quickly found out that the remote control was used to a minimal extent. The elderly people had a choice of 21 basic channels that they could watch; however, all four of them were regularly watching 4 channels or less. Apart from that, only the 'on button' was regularly used. The volume button was used occasionally. The first design suggestion for this user group was to make a very simple interface, with just a few (four) choices for channels, an 'on/off button' and a 'volume button.' Here, the affordances would be very clear, but so would the limitations. The question that presented a challenge was *how to take the step from the ease of use towards actual mastery* of an interface that could open for more possibilities if needed. The suggested design is shown on Fig. 3, where the possibility to choose more channels (the '*Flere kanaler* button' on the bottom right) or to perform some basic system functions (the '*Innstillinger* button', bottom left). A display of the time was also added.

Testing this new prototype brought another challenge: one of the participants had hand tremors and it was difficult for the participant to touch only the desired area since the buttons were too small. Adding some physics, so that more pressure needed to be applied by this participant than by other users on their interfaces was a way to resolve this challenge for the purposes of the project.

This is, however, not an easy task [13, 26, 28, 29]. A light, accidental touch should not activate a wrong channel. There are many ways to solve this problem, e.g., combining the touch with tangibles and/or sound.

Since the variation in the elderly population is so vast, a possible approach to a more general design solution could be through multimodality, configurability and personalization.

Figure 3. The second prototype implemented on an iPad [18].

C. *Video conferencing on the iPad.*

Our last two examples are student projects [30, 25] both focusing on the feeling of being alone, which is so common among elderly [12]. Both projects were concerned with providing easy access to video conferencing, in order to enable easy access to friends, family of caretakers. The solutions were developed on the iPad, using open source software whenever possible. The focus was on the closer understanding of touch functionality for elderly. Therefore, in case of the project [30], Fig. 4, a heat map was developed to show how wide is the range of places where elderly apply touch, while trying to press the green phone icon. This is contrasted to the heat map showing a group of students who had the same task. It is easy to see that, while the students are always on the target icon, the elderly touch everywhere, including empty spaces.

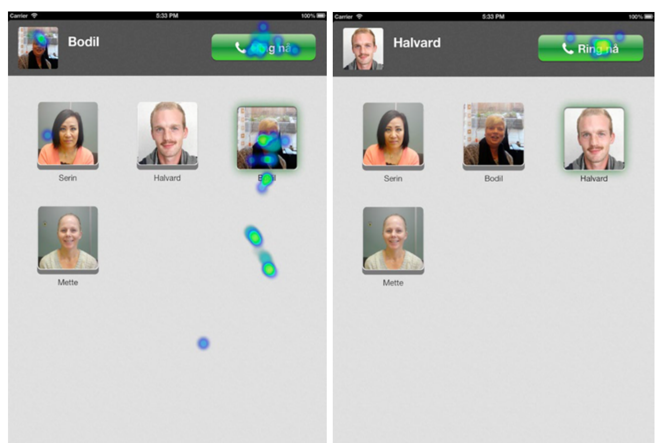


Figure 4. A heat map showing where elderly touch the screen when trying to reach the call button (left) vs. students (right) [31].

In [25], the students have focused on providing a large area where the touch may be applied; combining this with large icons similar to those of phone in Fig. 1; see Fig. 5.

Figure 5. Simple interface for calls [25].

III. Challenges in Design for and with Elderly

We face three major kinds of design challenges in the ‘technology@home’ project. The first is concerned with our choice of PD as design approach, the second stems from the loss of abilities that elderly people experience as they grow older, and the third is our focus on touch interaction.

A. *Participatory Design Challenges*

A challenge specific for PD with vulnerable users is how well they can represent themselves and have a say in the design process. Vulnerable users’ interests (e.g., children, disabled or elderly people) are often represented by those whose task is to work with and for them (case handlers, nurses, teachers, etc.) potentially leading to conflict where the two have differing interests. It is therefore of interest also in design of touch interfaces to explore ways in which less articulate groups of users can be given a stronger position as a PD participant. Thus, the first design challenge is concerned with arranging the participatory design process in ways that accounts for potential vulnerabilities and challenges that participants are meeting and at the same time enable active and creative engagement in the design process [3,16, 19, 20].

B. *Challenges from Loss of Abilities*

An important set of challenges in design for and with elderly people stems from their loss of physical and cognitive abilities. Such impairments often do not originate from one source only. In design, it is important to see how the elderly people themselves perceive their impairments and how these affect their everyday life [5]. For example, reduced vision and dementia are both frequent among those over 70 years. Clearly, both may affect a significant percentage of elderly population. But each elderly individual see him/herself differently and the way these two impairments play out in a person’s life would most likely differ from one person to another. It is understandable that because of the complexity of issues, designers often choose participants with only one impairment, such as dementia [20]. In-depth studies of the more complex composition of health issues normally present for elderly people suggests that technical solutions must be seen as parts of a system – an evolving system – and fit together. A basic infrastructure (like the ‘smart home’ package) can be a first step in the design of a set of technology tools that can be combined as needs change and appear.

C. *Touch Interface Challenges*

The third challenge concerns touch interaction and how this kind of interaction addresses bodily changes stemming from old age. Hornecker and Buur [14] introduce a conceptual framework for characterizing tangible interaction with four themes: 1) tangible manipulation, 2) spatial interaction, 3) embodied facilitation, and 4) expressive representation (Fig. 6). All four themes introduce limitations and possibilities based on the challenges posed by elderly users. For themes 1 and 2 the loss of abilities will directly limit the technical solutions, while new technologies offer new possibilities for themes 3 and 4.

Tangible Interaction			
Tangible Manipulation	**Spatial Interaction**	**Embodied Facilitation**	**Expressive Representation**
Haptic Direct Manipulation	Inhabited Space	Embodied Constraints	Representational Significance
	Configurable Materials		
Lightweight Interaction	Non-fragmented Visibility	Multiple Access Points	Externalization
	Full Body Interaction		
Isomorph Effects	Performative Action	Tailored Representations	Perceived Coupling

Figure 6. Tangible interaction framework [14, p. 440].

The tangible interaction theme introduces challenges concerned with haptic direct manipulation, where age-related changes constitute challenges of touch and grip. We need to explore whether the interpretation of lightweight interaction and isomorph effects is consistent for different age groups. Isomorph effects may turn out to be very important for people with cognitive impairment. The second theme, spatial interaction, seems particularly relevant for the home environment where the technical things should be integrated and fit in. Many elderly people arrange their home space in order to help them remember things (letters by the door, bills on the fridge, etc. [22]). The three remaining elements of this theme (non-fragmented visibility, full-body interaction and performative action) seem to not be particular for elderly users. The theme embodied facilitation concern how we move in space. Here, we focus on how the physical arrangement ease or delimit activities – and where the particular needs of elderly people are important to address. We are concerned with tailoring the representations to this particular user group and to offer multiple access points. The fourth theme concerns expressive representations: how the digital functions and data are represented physically. Here, we focus on the representational significance [24] conveying the state of the system to the user, the externalization as well as the perceived coupling between the physical object and the digital representation it embodies.

The *'technology@home'* will explore these themes from the perspective of the elderly user. Possibilities for tailoring the technology to the needs of each individual seems to be necessary for supporting elderly in their (technology-filled) homes. However, this is not possible to do neither practically nor economically. We will explore ways of customizing or individualizing a general solution as a easy-to-do part of the solution itself.

D. Implications for Design

Design grounded in the characteristics of elderly users may have to reconsider some design issues. One important issue is ethics. Ethical considerations concern the PD process and how the research and development is carried out [16] as well as the technical solutions resulting from the process [9]. Connected to the question about ethics is the announced aim to "improve quality of life" through technology – an aim that makes sense if we see technology as a support for and replacement of lost abilities. Mastery and autonomy of their own life is a basic element of social and emotional well-being. Home automation should be designed as something complementary to human care, not as a replacement. Interestingly, there are many initiatives to solve social problems and fear of loneliness with another series of technological solution falling under the category empathic design [2, 25, 31].

The major implication for design is that current touch screen designs rest on some assumptions that are often not present for elderly users: they are sometimes not able to see what to do on a small screen, nor are they able to push a small area in a way which is not too hard or soft, or too long or short. The elderly users' abilities are different from person to person and they change over time. It therefore seems necessary to include in the design an automatic customization procedure where the system gets to know the particular user and her/his particular way of touching the screen.

IV. Concluding Remarks

Designing with and for elderly carries a complex set of issues ranging from ethical considerations and methodological challenges to the choice of interaction styles and modalities. Many challenges are related to their vulnerability, decline in physical and cognitive abilities and diversity in manifestation of this decline among elderly. A smart home can be viewed as ecology of devices, potentially interacting with each other and/or with the outside world. If the elderly are to be supported by these devices, they need to be willing to, and know how to use them for their own well-being. Everything they use needs to be customizable for them, as well as adaptable to their changing abilities. Mastery and autonomy are among basic ingredients needed for the feeling of well-being. If the technology can help elderly to accomplish greater degree of independence through mastery and autonomy, its promise and opportunity is fulfilled. To approach that goal, we believe PD is a good methodological approach for evaluation of use of existing solutions and designing new solutions for elderly. When it comes to the naturalness and ease of touch-screens for elderly, we cannot conclude that they are an optimal choice. However, with customization and adaptation strategies, they may become a better match. This is also a direction for our future research.

Acknowledgment

The authors are grateful to Oslo municipality, elderly who participated in the study and students participating in projects involving elderly users.

References

[1] Ambient Assisted Living joint programe http://www.aal-europe.eu/ [retreived: January, 2013].

[2] M. Asboe, J. Fernandes, and E. Grönvall, "C-HUB: a communication and network platform targeting the Generation Plus and their social and care networks," *DHRS2010*, pp. 7, 2010.

[3] T. Bratteteig, and I. Wagner, Beyond the smart home. Designing multimodal environments to support mastery design, *4S conference*, Denmark, September 2012.

[4] B. Challis, Tactile Interaction. In Soegaard, et al. eds. *Encyclopedia of Human-Computer Interaction* (2012). Available online at http://www.interactiondesign.org/ encyclopedia/tactile_interaction.html [retreived: January, 2013].

[5] K. Cheverst, K. Clarke, G. Dewsbury, T. Hemmings, S. Kember, T. Rodden, and M. Rouncefield, "Designing assistive technologies for medication regimes in care settings," *Universal Access in the Information Society*, vol. 2, no. 3, pp. 235–242, 2003.

[6] A. Gasparini, A. Culén, Tablet PCs – An Assistive Technology for Students with Reading Difficulties? ACHI 2012, Fifth International Conference on Advances in Computer-Human Interactions (2012), pp. 28–34.

[7] A. Dickinson, J. Arnott, and S. Prior, "Methods for HCI research with older people". *Behavior & Information Technology, 26, 4,* 2007, pp. 343-352.

[8] W. K. Edwards and R. E. Grinter, "At Home with Ubiquitous Computing: Seven Challenges," in *Proceedings of the 3rd international conference on Ubiquitous Computing*, London, UK, UK, 2001, pp. 256–272.

[9] A. Essén, The two facets of electronic care surveillance: an exploration of the views of older people who live with monitoring devices". Social Science & medicine, 2008, 67 (1), pp. 128-136.

[10] S. M. Golant and J. Hyde, Eds., *The Assisted Living Residence: A Vision for the Future*, 1st ed. The Johns Hopkins University Press, 2008.

[11] J. Grudin, "Groupware and social dynamics: eight challenges for developers," *Commun. ACM*, vol. 37, no. 1, 1994, pp. 92–105.

[12] Helsedirektoratet. Welfare technology. Report on implementation of welfare technologies in the municipal helath and care services 2013-2030, 2012.

[13] T. Hirsch, J. Forlizzi, E. Hyder, J. Goetz, C. Kurtz, and J. Stroback, "The ELDer project: social, emotional, and environmental factors in the design of eldercare

technologies," in *Proceedings on the 2000 conference on Universal Usability*, New York, NY, USA, 2000, pp. 72–79.

[14] E. Hornecker and J. Buur, "Getting a grip on tangible interaction: a framework on physical space and social interaction," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006, pp. 437–446.

[15] D. Howthorn, "Possible implications of aging for interface designers." *Interacting with Computers. 12*, 2000, pp. 507 – 528.

[16] S. Jesper, and T. Robertson. *Routledge Handbook of Participatory Design*. Routledge International, 2012.

[17] M. Kobayashi, A. Hiyama, T. Miura, C. Asakawa, M. Hirose, and T. Ifukube, "Elderly User Evaluation of Mobile Touchscreen Interactions," in *Human-Computer Interaction – INTERACT 2011*, vol. 6946, P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, and M. Winckler, Eds. Springer Berlin / Heidelberg, 2011, pp. 83–99.

[18] K. Larsen, G. Gahnstrøm, K. Matyja and J. P. Nilssen. Student project "Team Old School": http://www.uio.no/studier/emner/matnat/ifi/INF2260/h11/prosjekter/inf2260-team-old-school-rapport-v2.pdf [retreived: January, 2013].

[19] S. Lindsay, D. Jackson, G. Schofield, and P. Olivier, "Engaging older people using participatory design," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, New York, NY, USA, 2012, pp. 1199–1208.

[20] S. Lindsay, K. Brittain, D. Jackson, C. Ladha, K. Ladha, and P. Olivier, "Empathy, participatory design and people with dementia," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, New York, NY, USA, 2012, pp. 521–530.

[21] D. López and M. Domènech, "Embodying autonomy in a Home Telecare Service," *The Sociological Review*, vol. 56, 2008, pp. 181–195.

[22] L. Palen & S.Aaløkke Ballegaard. Of pill boxes and piano benches: "home-made" methods for managing medication. CSCW, 2006, pp. 79-88.

[23] M. Rauhala and P. Topo, "Independent living, technology and ethics," *Technology and Disability*, vol. 15, no. 3, 2003, pp. 205–214.

[24] B. Ullmer and H. Ishii. Emerging frameworks for tangible user interfaces. IBM Systems Journal 39 (3-4), 2000, pp. 915-931.

[25] A. Voje, H. Eggen, T. Gabrielsen, T. Vitikka and L. Bech Student project "Kamfer": http://www.uio.no/studier/emner/matnat/ifi/INF2260/h12/projects/elderly/Kamfer/ [retreived: January, 2013].

[26] C. Wacharamanotham, J. Hurtmanns, A. Mertens, M. Kronenbuerger, C. Schlick, and J. Borchers, "Evaluating swabbing: a touchscreen input method for elderly users with tremor," in *Proceedings of the 2011 annual conference on Human factors in computing systems*, 2011, pp. 623–626.

[27] R. Wang and F. Quek, "Touch & talk: contextualizing remote touch for affective interaction," in *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, 2010, pp. 13–20.

[28] F. Werner, K. Werner, and J. Oberzaucher, "Tablets for Seniors – An Evaluation of a Current Model (iPad)," in *Ambient Assisted Living*, R. Wichert and B. Eberhardt, Eds. Springer Berlin Heidelberg, 2012, pp. 177–184.

[29] X.J. Zhao, T. Plocher, and L. Kiff, "Touch screen user interfaces for older adults: button size and spacing". In: *HCI International 2007,* Beijing, China, July 22–27, 2007.

[30] V. S. Åmdal, K. Klette, H, L. Simonsen and S. Ahmedov "Selskap": http://www.uio.no/studier/emner/matnat/ifi/INF2260/h12/projects/elderly/Elderly-aid-app/ [retreived: January, 2013].

Basic Study for New Assistive System Based on Brain Activity during Car Driving

Shunji Shimizu / Tokyo University of Science, Suwa
Department of Electric Systems Engineering
Chino-city, Japan
shun@rs.suwa.tus.ac.jp

Hiroaki Inoue/ Tokyo University of Science, Suwa
Research course of Engineering/Management
Chino-city, Japan
jgh12701@ed.suwa.tus.ac.jp

Hiroyuki Nara/ Hokkaido University
Graduate School of Information Science and Technology
Sapporo-city, Japan
nara@ssc.ssi.ist.hokudai.ac.jp

Noboru Takahashi/ Tokyo University of Science, Suwa
Research course of Engineering/Management
Chino-city, Japan
srl@rs.suwa.tus.ac.jp

Fumikazu Miwakeichi/ The Institute of Statistical Mathematics
Spatial and Time Series Modeling Group
Tachikawa-city, Japan
miwake1@ism.ac.jp

Nobuhide Hirai / Jichi Medical University
Department of Psychiatry
Shimotsuke-city, Japan City
nobu@nobu.com

Senichiro Kikuchi / Jichi Medical University
Department of Psychiatry
Shimotsuke-city, Japan
skikuchi@jichi.ac.jp

Eiju Watanabe/ Jichi Medical University
Department of Neurosrugery
Shimotsuke-city, Japan
eiju-ind@umin.ac.jp

Satoshi Kato/ Jichi Medical University
Department of Psychiatry
Shimotsuke-city, Japan
psykato@jichi.ac.jp

***Abstract*— Final purpose in this research is to contribute to developing of assistive robot and apparatus. Recently, there is a pressing need to develop a new system which assists and acts for car driving and wheelchair for the elderly as the population grows older. In terms of developing a new system, it is thought that it is important to examine behaviors as well as spatial recognition. Therefore, experiments have been performed for an examination of human spatial perceptions, especially right and left recognition, during car driving using NIRS. In previous research, it has been documented that there were significant differences at dorsolateral prefrontal cortex at left hemisphere during virtual driving task and actual driving.**

In this paper, brain activity during car driving was measured and detailed analysis was performed by segmental brain activity during car driving on the basis of subjects' motion. So, we report the relationship between brain activity and movement concerned with perception during driving in this paper.

Keywords-brain information processing during driving task; spatial cognitive task;determining direction; NIRS.

I. INTRODUCTION

Human movements change relative to his environment. Nevertheless, he/she recognizes a new location and decides what behavior to take. It is important to analyze the human spatial perception for developing autonomous robots or automatic driving.

The relation of the theta brain waves to the human spatial perception was discussed in [1][2]. When humans perceive space, for example, try to decide the next action in a maze, the theta brain waves saliently appear. This means we have a searching behavior to find a goal at an unknown maze. From the side of human navigation Maguire et al. measured the brain activations using complex virtual reality town [3]. But every task is notional and the particulars about the mechanism that enables humans to perceive space and direction are yet unknown. Also, Brain activities concerned with cognitive tasks during car driving have been examined. For example, there was a report about brain activity when disturbances were given to subjects who manipulated driving simulator. Also, power spectrums increased in beta and theta bands [4]. However, there is little report on the relationship among right and left perception and driving task.

 ISBN: 978-1-61208-250-9

So, we performed experiments in which perception tasks were required during virtual car driving using NIRS(Near Infrared Spectroscopy). From experimental results, there were significant differences at dorsolateral prefrontal cortex in left hemisphere via one-sample t-test when subjects watched driving movie and moving their hand in circles as if handling a steering wheel [5].

In addition, we conducted experiments in real space, which were performed by taking f-NIRS in the car, and measured the brain activity during actual driving. A purpose in this experiment was to measure and analyze the brain activity during actual driving to compare results between virtual and actual results. As a result, there were significant differences at similar region [6][7]. In addition, we measured the brain activity of frontal lobe, which is related to behavioral decision-making, during car driving in different experimental design from previous one to verify previous results [8][9].

It is well known that higher order processing is done such as memory, judgment, reasoning, etc in the frontal lobe [10]. We tried to grasp the mechanism of information processing of the brain by analyzing data about human brain activity during car driving. Also, the goal of this study is to find a way to apply this result to new assist system.

So, with the aim of increasing number of subjects and examining more closely the brain activity concerned with spatial perception and direction determination during car driving, we performed additional experiments.

In this time, the brain activity of same lobe with human spatial perception and direction determination was discussed on the basis on changing direction of the gaze and starting to turn the steering wheel. Furthermore, we examined the mechanism of information processing of the brain and human spatial perception during car driving.

II. Experiment

A. Brain activity on virtual driving

1) Brain activity on driving movie is shown

The subjects for this experiment were eight males who were right handed. They were asked to read and sign an informed consent regarding the experiment.

An NIRS (Hitachi Medical Corp ETG-100) with 24 channels (sampling frequency 10 Hz) was used to record the density of oxygenated hemoglobin (oxy hemoglobin) and deoxygenated hemoglobin (de-oxy hemoglobin) in the frontal cortex area.

The movie is included two scenes at a T-junction in which it must be decided either to turn to the right or left. In the second scene, there is a road sign with directions. We used nine kinds of movies in about one minutes.Before showing the movie, subjects were given directions to turn to the right or left at the first T-junction. They were also taught the place which was on the road sign at the second T-Junction. They had to decide the direction when they looked

Fig. 1. Recorded movie during measurement

at the road sign. They were asked to push a button when they realized the direction in which they were to turn.

2) Brain activity on handling motion

In this experiment, measuring was performed by f-NIRS(Functional Near Infrared Spectroscopy), made by SHIMADZU Co. Ltd with 44ch. Five subjects were healthy males in their 20s, right handed with a good driving history. They were asked to read and sign an informed consent regarding the experiment.

The subject was asked to perform simulated car driving, moving their hand in circles as if using a steering wheel. A PC mouse on the table was used to simulate handling a wheel, and NIRS (near-infrared spectroscopy) to monitor oxygen content change in the subjects' brain. NIRS irradiation was performed to measure brain activities when the subject sitting on a chair make a drawing circle line of the right/left hand 1) clockwise, and 2) counterclockwise. The part of measurement was the frontal lobe. The subject was asked to draw on the table a circle 30 cm in diameter five times consecutively, spending four seconds per a circle. The time design was rest (10 seconds at least) – task (20 seconds) – rest (10 seconds) - close rest.

B. Brain activity during actual car driving

1) Brain activity during actual car driving

In general roads, experiments were performed by taking f-NIRS in the car, and measuring the brain activity when car driven by subjects was went through two different intersections. Six subjects were a healthy male in their 20s, right handed with a good driving history. They were asked to read and sign an informed consent regarding the experiment. In all experiments, measuring was performed by f-NIRS (Functional Near Infrared Spectroscopy), made by SHIMADZU Co. Ltd [11].

Subjects took a rest during 10 seconds at least with their eye close before driving task and they drove a car during about 600 seconds. Finally, subject closed their eyes for 10 seconds again after task. Then, the brain activity was recorded from the first eyes-closed rest to the last eyes

Subjects were given directions to turn to the right or left at the first T-junction during driving task. They were also taught the place which was on the road sign at the second T-junction. And, they were given the place where they have to

 ISBN: 978-1-61208-250-9

Fig. 2. Sample of first T-Junction

Fig. 3. Sample of second T-Junction

go to. So, they had to decide the direction when they looked at the road sign.

A trigger pulse was emitted on stop lines at T-Junctions to use as a measuring stick for the analysis. Also, we recorded movie during the experiment from a car with a video camera aimed toward the direction of movement (Fig. 1). Recorded movies were used to exempt measurement result including disturbances, such as foot passengers and oncoming cars, from analysis. Figure.2 and figure.3 shows one sample of T-junction.

2) Verification Experiment

To conduct verification for experimental results in previous experiment, we performed additional experiment which was achieved in a similar way.

In this experiment, experimental course was different from previous one. While previous one was included two T-junctions in which there was road sign at second one and not at first one per a measurement, there were multiple T-junctions. Three were 5 T-junction without road sign and 4 T-junctions with road sing.

Subjects were twelve males who were all right-handed. They drove a car during about 20 minutes after a rest during 10 seconds at least with their eyes close. Subjects were enlightened about turning direction and the place on which road signs was at T-junction during measurement. And, they arbitrarily decided the direction to turn when they confirmed road signs. Also, a trigger pulse was emitted in the same way.

3) Detailed analysis based on driving behavior

In this analysis, movies aimed toward the direction of movement as well as ones aimed subject movement like ocular motion and arm movement were recorded. This is to analyze brain activity using ocular motion in looking at road signs as a trigger. In previous research we performed, stop line at T-junction was used as a trigger. But, brain activity in T-junction involved movement task such as turning steering wheel, changing neck direction, hitting the brake. So, it is thought that brain activity derived from cognitive tasks was overwritten with brain activity due to movement tasks. Therefore, we tried to analyze brain activity on the basis of ocular motions to examine significant differences with cognitive tasks.

III. EXPERIMENTAL RESULT

A. Brain activity on virtual driving

1) Brain activity on driving movie is shown

On the whole, the variation in de-oxy hemoglobin was smaller than in the oxy hemoglobin. However, there was a great increase in channel 18(around #10 area of the dorsolateral prefrontal cortex of the right hemisphere). This might be the variation based on the spatial perceptions

Next, differences were investigated concerning the subject's brain activity. As the first case, it was when the vision was directed after having been told the direction. As the Second, it was when the vision was directed after having been decided the direction under the road sign.

Here, d1 and d2 were defined to analyze measurement data. d1 is the variation of hemoglobin turning of one second at the first T-junction. And d2 is variation of hemoglobin at the second one. From the measurement result, d1 and d2, all of the 269 times of each subject, there were significant differences in oxy hemoglobin 3ch. ($p<0.02$: paired t test) and 20ch. ($p<0.03$) using NIRS. These regions were corresponded to around #46 area of the dorsolateral prefrontal cortex of the left hemisphere and around #10 area of the dorsolateral prefrontal cortex of the right hemisphere, respectively.

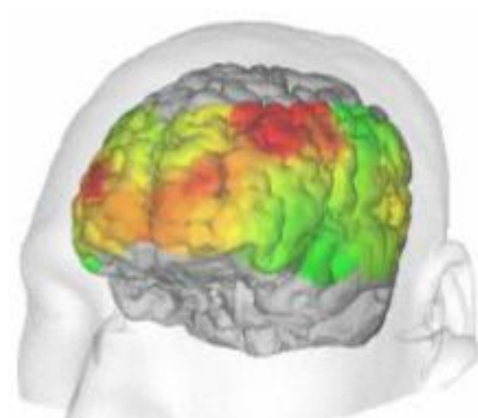


Fig. 4. Brain activity (clockwise)

 ISBN: 978-1-61208-250-9

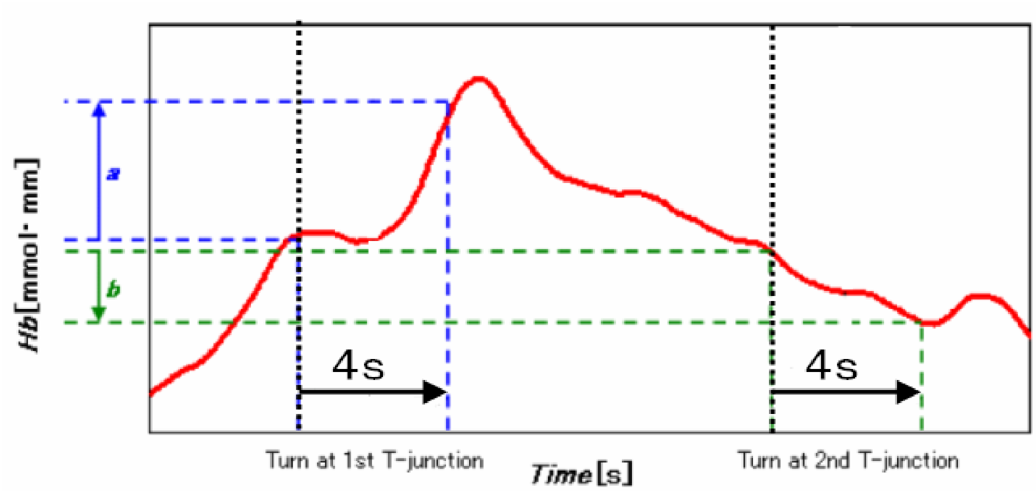


Fig. 5 Analysis method

Subjects pushed a button before turning at the second T-junction, so it influenced brain activities. The possibility of a correlation between d2 and the time until the movie was turned at the second T-junction after each subject pushed a button was investigated. Each correlation coefficient of hemoglobin channel was calculated. There was significant difference at only de-oxy hemoglobin 10ch(around #10 area of the dorsolateral prefrontal cortex of the right hemisphere) using paired t test. In only this result, the relationship between pushing a button and d2 cannot be judged.

2) Brain activity on handling motion

During the motion, the increase of oxy hemoglobin density of the brain was found in all subjects. The different regions of the brain were observed to be active, depending on the individual. The subjects were to be observed 1) on starting, and 2) 3-5 seconds after starting moving their 3) right hand 4) left hand 5)clockwise 6)counterclockwise. Although some individual variation existed, the result showed the significant differences and some characteristic patterns. The obtained patterns are shown as follows. Regardless of 1), 2), 3) and 4) above, the change in the oxy hemoglobin density of the brain was seen within the significant difference level 5% or less in the three individuals out of all five subjects. The part was the adjacent part both of left pre-motor area and of left prefrontal cortex. Especially, in the adjacent part of prefrontal cortex a number of significant differences were seen among in four out of five subjects. Next more emphasis was put on the rotation direction: 5) clockwise or 6) counterclockwise. No large density change was found in the brain with all the subjects employing 6). But, the significant difference was seen in four out of five subjects employing 5) (Fig.4). It is well known that in the outside prefrontal cortex higher order processing is done such as of behavior control. It is inferred that the pre-motor area was activated when the subjects moved the hand in the way stated above because the pre-motor area is responsible for behavior control, for transforming visual information, and for generating neural impulses controlling.

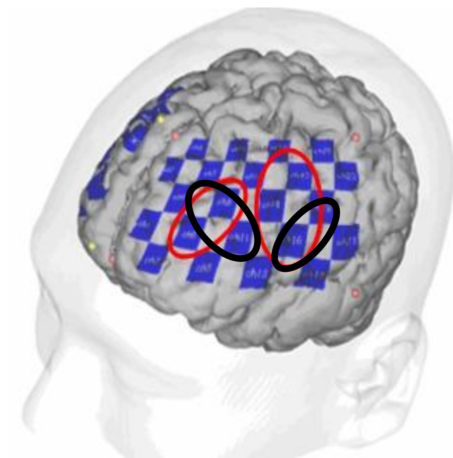


Fig. 6 Significant differences at the turn left

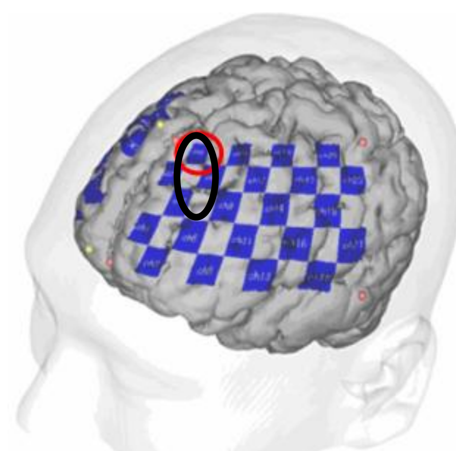


Fig. 7. Significant differences at the turn right

B. Brain activity during actual car driving

1) Brain activity during actual car driving

At the first, Hb-oxy was increased in overall frontal lobe after start of operation. This tendency was common among subjects. After that, Hb-oxy was decreased as subjects adjusted to driving the car. This meant that the brain activity changed from collective to local activities.

In this experiment, being considered time as zero when experimental vehicle reached stop line at T-junction. The analysis was performed one-sample t-test using a and b within the significant difference level 5% or less between zero and about four seconds (Fig.5). Here, a is the variation of hemoglobin turning of one second at the first T-junction. And b is variation of hemoglobin at the second one. As the results, there were significant differences around #46 area of the dorsolateral prefrontal cortex and the premotor area of the left hemisphere brain in turning left(Fig. 6:red). Also, there were significant differences #9 of the dorsolateral prefrontal cortex of the left hemisphere brain at the turn right (Fig. 7: red).

2) Verification Experiment

Various tendencies among individuals were observed in comparison with results in B. However, there were tendency that oxy-Hb was increased when car turned left or right at T-junctions and oxy-Hb was decreased during going straight

Analysis method was the same as previous one. Though Gaps were shown regions at which there were significant differences, there were significant differences in common region, too (Fig. 6,7: black). In the analysis, measurement results including disturbance at T-junctions were excluded as analysis object.

3) Detailed analysis based on driving behavior

The analysis was performed one-sample t-test within the significant difference level 5% or less between brain activity before and after looking at road sign. Each of sample data for analysis was 1 second. Also, analysis was performed with respect to each direction which subject had to go at next T-junction. As a consequence of analysis, there were significant differences at interior front gyrus of flontal lobe of left hemisphere without reference of direction (Fig. 8 and Fig. 9).

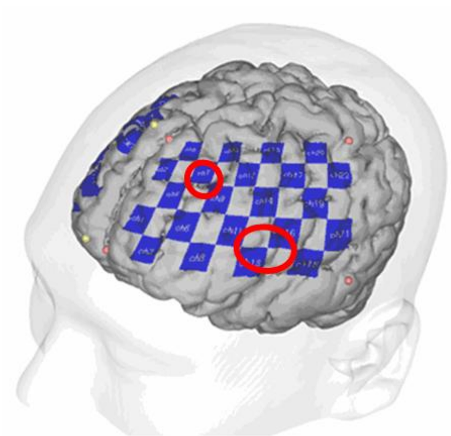


Fig. 8. Significant difference at left direction of road sign

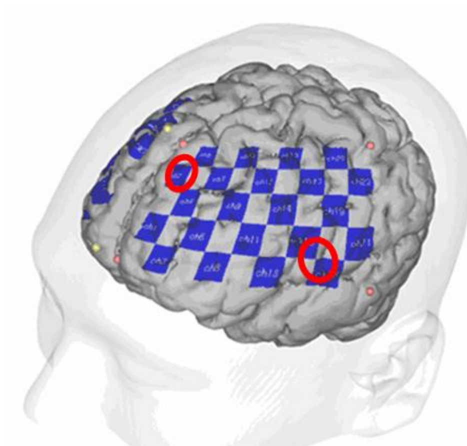


Fig. 9. Significant difference at right direction of road sign

IV. CONCLUSION AND FUTURE WORK

The hemoglobin density change of the human subjects' frontal lobe was partly observed in the experiments we designed, where three kinds of tasks were performed to analyze human brain activity from the view point of spatial perception.

The NIRS measures of hemoglobin variation in the channels suggested that human behavioral decision-making of different types could cause different brain activities as we saw in the tasks: 1) taking a given direction at the first T-junction, 2) taking a self-chosen direction on a road sign at the second T-junction and 3) turning the wheel or not. Some significant differences (paired t test) on NIRS's oxy-hemoglobin and less interrelated results between "pushing a button" and brain activity at the second T-junction are obtained.

Furthermore, experimental results indicated that with the subjects moving their hand in circle, regardless of right or left, 1) the same response was observed in the prefrontal cortex and premotor area, and 2) different patterns of brain activities generated by moving either hand clockwise or counterclockwise.

The regions observed were only those with the 5% and less significance level. Possible extensions could be applied to other regions with the 10% and less significance level for the future study. With a larger number of subjects, brain activity patterns need to be made clear. In addition, it is thought to take particular note of participation concerning working memory when car is driven.

Furthermore, it was found that there were significant differences around #44-45 area. It is well known that this region is corresponding to language area. So, it is thought that subjects look at road map to determine direction that they have to go according to word described in road sign.

From results of these experiments, there was significant difference around working memory. So, experiments focusing on relationship turning wheel and working memory will be performed. On the other hand, experiments as to actual driving were required a broad range of perception and information processing. Especially, subjects had to determine behaves depending on various information at T-junctions, that is, the color of the traffic light, presence or absence foot passengers and so on. And so, we plan to perform more static experiments. we attention to differences on the basis of turning direction and dominant hand. In addition, we will conduct the experiments in which subjects were narrowed down to left-handedness. Furthermore, researches into other human brain activities than spatial perception are to be necessary with accumulated data from fMRI (functional magnetic resonance imaging), EEG (Electroencephalogram), etc.

When compared virtual result to actual ones, there were significant differences around #46 area in both experiments, which were performed in virtual and actual condition, as a common result. It is thought that this result is due to activities of working memory because subjects must to recall memories of movements required for car driving and turning steering wheel. Conversely, there were significant differences around #10 in virtual experiments and around premotor area in actual driving, respectively. In the virtual case, it is thought to result from inhabitation of task without movement. In the actual case, subjects had to perceive space information in real time. So, it is considered that there were significant differences around premotor area because they always ready up to manipulate steering wheel.

As a future plan, we aim to apply these results to assistive human interface. As a matter of course, we plan to performed additional experiments including the verification of these results. And final purpose is to develop a new system for manipulating wheelchair and information presentation system to assist recognition of information including spatial one during car driving.

REFERENCES

[1] M.J. Kahana, R. Sekuler, J.B. Caplan, M. Kirschen, and J.R. Madsen: “Human theta oscillations exhibit task dependence during virtual maze navigation.”, Nature, 1999, 399, pp. 781-784.

[2] N. Nishiyama and Y. Yamaguchi: "Human EEG theta in the spatial recognition task", Proceedings of 5th World Multiconf. On Systemics, Cybernetics and Informatics (SCI 2001), Proc. 7th Int. Conf. on Information Systems, Analysis and Synthesis (ISAS 2001), pp. 497-500 (2001).

[3] E.A. Maguire, N. Burgess, J.G. Donnett, R.S.J. Frackowiak, C.D. Frith, and J.O’ Keefe: "Knowing Where and Getting There: A Human Navigation Network," vol. 280 Science 8/may/1998.

[4] Chin-Teng Lin, Shi-An Chen, Tien-Ting Chiu, Hong-Zhang Lin, and Li-Wei Ko: "Spatial and temporal EEG dynamics of dual-task driving performance." Journal of NeuroEngineering and Rehabilitation, vol. 8-11, 2011

[5] S. Shimizu, N. Hirai, F. Miwakeichi, and et al: "Fundamental Study for Relationship between Cognitive task and Brain Activity during Car Driving," Proc. the 13th International Conference on Human-Computer Interaction, (San Diego, CA, USA, 2009), Springer Berlin / Heidelberg, 434-440.

[6] N. Takahashi, S. Shimizu,Y. Hirata, H. Nara, F. Miwakeichi, N. Hirai, S. Kikuchi, E. Watanabe, and S. Kato: "Fundamental Study for a New Assistive System during Car Driving," Proc. International Conference on Robotics and Biomimetics, 2010, China.

[7] N. Takahashi, S. Shimizu, Y. Hirata, H. Nara, H. Inoue, N. Hirai, S. Kikuchi, E. Watanabe, and S. Kato,“Basic study of Analysis of Human Brain Activities during Car Driving,” the 14th International Conferrence on Human-Computer Interaction, 2011, Orlando, Florida, USA.

[8] S. Shimizu, N. Takahashi, H. Nara, H. Inoue, and Y. Hirata ,“Fundamental Study for Human Brain Activity Based on the Spatial Cognitive Task,” the 2011 Internatinal Conference on Brain Informatics-BI 2011, China.

[9] S. Shimizu, H. Nara, N. Takahashi, H. Inoue and, Y. Hirata,“Basic Study for Human Brain Activity Based on the Spatial Cognitive Task,” The Third International Conference on Advanced Cognitive Techonologies and Applications, 2011, Italy.

[10] J. Cockburn: "Task interruption in prospective memory: “A frontal lobe function?." Cortex, vol. 31, 1995, pp. 87- 97.

[11] E. Watanabe, Y. Yamashita, Y. Ito and, H. Koizumi, ”Non-invasive functional mapping with multi-channel near infra-red spectroscopic topography in humans,” Heurosci Lett 1996, Feb 16, 205(1), 41-4.

Proposal of an Automobile Driving Interface Using Gesture Operation for Disabled People

Yoshitoshi Murata, Kazuhiro Yoshida
Faculty of Software and Information Science
Iwate Prefectural University
Takizawa, Japan
y-murata@iwate-pu.ac.jp, kyoshida@ipu-office.iwate-pu.ac.jp

Kazuhiro Suzuki, Daisuke Takahashi
JFP Inc.
Morioka, Japan
ss@jfp.co.jp, takahashi@jfp.co.jp

***Abstract*— A steering operation interface has been designed for disabled people that uses gesture operation. It incorporates both non-linear and semi-automatic steering control. Experiments using a gyro sensor and a driving simulator demonstrated that the driving operation is close to that achieved with conventional steering wheel operation. Sufficient practice in using the proposed interface should therefore enable a user to achieve steering control closer to that achieved with a steering wheel.**

Keywords-automobile driving interface; disabled people; gyro sensor; gesture operation; body part operation

I. Introduction

Disabled people generally want to stand on their own two feet, and achieving mobility is an important step in doing this. One way to enhance mobility is through the use of an automobile to which a driving-assistance device has been attached. However, there has been a lack of development of auxiliary devices and systems that would enable disabled people, especially people with arm and wrist disabilities, to drive a car. For example, the first such system, the original of Honda's Franz system [1], was developed in the 1960s. With this system, a car is operated with only the feet. Since the steering wheel is turned by pumping the pedals, its operation is not intuitive.

With the system developed by Wada and Kameda, the steering wheel is controlled with a joystick, and the brake and accelerator are controlled with another joystick [2][3]. This system has aided many disabled people, but strength is needed to operate the joysticks. Moreover, the levers onto which the joysticks are fixed have to be customized for the hand positions of each user.

In any case, mechanical devices such as these lack flexibility and have to be customized for the user. Hence, they are inherently expensive.

The on-going shift from hydraulic to electronic driving interface systems (steering, braking, etc.) means that systems combining computer chips with sensors can now be used to easily control these driving interfaces. Candidate sensors include the KINECT sensor, a gyro sensor and so on.

We have designed a steering operation interface for disabled people that is operated by gestures, developed a prototype control device that uses a gyro sensor, and evaluated it by using a driving simulator as the initial stage of our research.

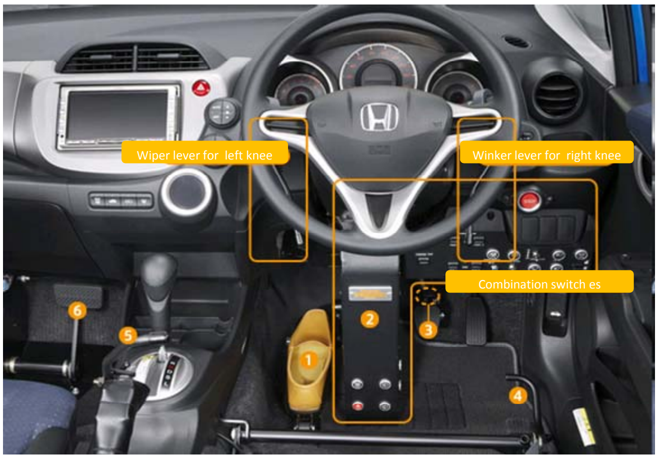
After discussing related work in Section II, we describe the gestures, i.e., body-part movements assigned to the various functions in Section III, the driving simulator we developed to evaluate our proposed driving interface in Section IV, and the experimental evaluation we conducted in Section V. The key points are summarized and future work is mentioned in Section VI.

Funding for this work was received from the Iwate Strategic Research Foundation.

II. Related Works

A. Driving Interface for Disabled people

The Franz system used by Honda was aimed at people who have difficulty moving their arms and/or hands. The user operates a car using only his or her feet [1]. It was originally implemented in the Honda Civic in 1982, creating the first



(1) Steering pedal
(2) Steering box
(3) Brake lock button
(4) Selection bar for feet
(5) Side brake for knee
(6) Sub-brake for exercise

Figure 1. Honda's Franz system

 ISBN: 978-1-61208-250-9

vehicle for disabled people in Japan. It has now been implemented in the Honda Fit.

The steering wheel is turned right or left by pumping a steering pedal (see Fig. 1). The transmission is shifted into drive by lifting the selection bar, into reverse by pushing it down, and into park by pushing it further down. The turn signals and windshield wipers are operated by turning levers using the right and left knees. Power windows, lights, and so on are controlled by flipping switches up or down using the right foot or knee.

Wada and Kameda developed a joystick car driving interface for wheelchair users. In the initial version [2][3], the steering, braking, and acceleration were controlled with one joystick. In the latest version, two joysticks are used, as shown in Fig. 2. The right joystick controls the steering, and the left one controls the acceleration and braking. The relationship between the angle of the steering wheel and the angle of the joystick is a polyline, as shown in Fig. 3. This means that a driver can sensitively control the steering wheel around the neutral position and can turn the wheel quickly when making a wide turn. People who can freely move their hands can drive an automobile using this device.

In any case, such mechanical devices must be customized to fit the user's disability and physical form.

B. Sensors for gesturing

Several driving interfaces using the KINECT sensor have been developed. With the "Air Driving" interface developed by Forum8, a user can drive a virtual car in a simulated world by moving his or her hands and feet in front of the sensor [4]. Since there must be at least 50 cm between the sensor and the gesturing body part, it cannot be used in an actual car. Rahman et al. developed a car audio operation interface that uses a KINECT sensor [5]. Although this interface has been demonstrated in an actual car, its use as a driving interface (steering, braking, etc.) has not been examined.

Döring et al. developed a multi-touch steering wheel that can control not only the steering but also the car audio [6]. However, users with an arm disability have trouble operating it.

Other examples of using an acceleration sensor and/or a gyro sensor as a gesture operation interface include video games and home appliance remote controls [7].

III. GESTURES (BODY-PART MOVEMENTS) FOR OPERATION

A. Operating functions

Door open/close, window open/close, wiper on/off, and turn signal on/off functions are needed to drive an automobile in addition to the basic operations of steering, braking and accelerating. Moreover, since automobiles typically have an audio system, a navigation system, and a climate control system, a driver should be able to operate these systems as well. Other than for the basic operating functions, a fine degree of control is not needed for the operating functions—they can generally be controlled by flipping a switch, as in Honda's Franz system. Moreover, voice-command control systems like that used for Samsung's SMART TV [8] could also be used.

Of the basic operations requiring a fine degree of real-time control (steering, braking and accelerating), we focused on steering, which requires the finest degree of control. Our research results should be easily transferable to braking and accelerating.

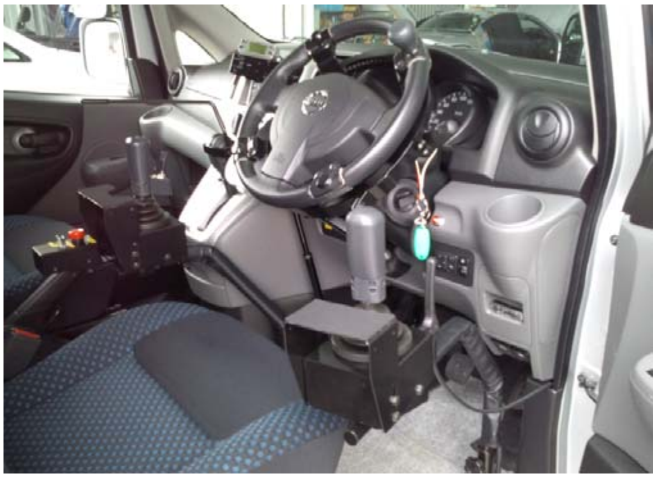


Figure 2. Wada and Kameda's joystick driving interface

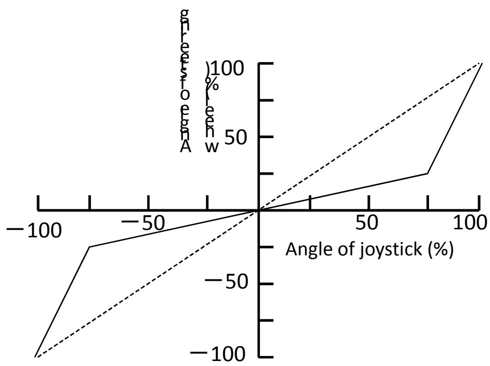


Figure 3. Relationship between angle of joystick and angle of steering wheel

B. Steering operation requirements

Steering an automobile by moving a body part should produce the same results as steering by turning the wheel by hand. Given this basic requirement, we derived several specific requirements.

1) The automobile should be steerable within ± several ten degrees from the neutral position.

- *There should be a fine degree of steering control around the neutral position.*
- *Steering should be quick when making a wide turn.*

2) The driver should be able to keep the vehicle within the lane on both straightaways and curves of various radii at a normal driving speed.

3) The driver should be able to drive stably, not zigzag, on straightaways

4) The driver should be able to traverse a curve while keeping the steering wheel at a position fixed immediately before entering the curve and then exit the curve into a straightaway by gradually returning the steering wheel to the neutral position.

C. Candidate sensors or devices

We considered four types of sensors or devices for controlling steering.

- KINECT sensor
- Video camera
- Rotary encoder
- Gyro sensor

Using a KINECT sensor or video camera is problematic because the unit has to be attached to the car, and the locations for possible attachment are limited. Moreover, as mentioned above, there must be at least 50 cm between a KINECT sensor and the gesturing body part, which greatly limits the possible attachment locations.

A rotary encoder requires the use of a mechanical adapter to measure the joint angle of a finger, an elbow, an ankle, and so on.

Since a gyro sensor is affected by not only the joint angle but also the vehicle motion, it must be attached to the vehicle to eliminate this effect. Moreover, a gyro sensor has a drift error that increases cumulatively. Since it is very difficult to remove the cumulative error completely, the measurement angle has to be reset using reliable data measured by another means.

A strain gauge does not have a drift error and is not affected by vehicle motion, so it should be better suited for measuring joint angles than a gyro sensor. We plan to investigate its usefulness in future work.

In this work, we investigated the use of a gyro sensor (ATR Promotion WAA-010) as well as a joint angle such as a finger, an elbow, or an ankle for driving a car.

D. Steering control

The steering wheel in an actual automobile can be turned about three complete revolutions from wheel lock to wheel lock (about 1080°). In contrast, the movable angle of a joint angle is about 20–90°, much less than that of a steering wheel. Hence, it is impossible to control the steering with a joint angle the same as is done with a steering wheel.

We thus introduce non-linear steering control and semi-automatic steering control. The direct operation angle and automatic steering angle are determined as shown in Fig. 4, which illustrates steering control using a foot and ankle. A driver operates the steering using the non-linear steering control within the direct operation angle. Whereas Wada and Kameda used a polyline function for their joystick steering control, we use a non-linear function ($y = x^n$). We set n = 3 on the basis of our experimental results, which are described in Section V. The steering angle increases automatically when it is beyond the direct operation angle. The rate of increase depends on the speed of the car: the faster the car, the lower the rate. The driver can stop a further increase in the steering angle by lifting his or her toes (about 20° for the case illustrated in Fig. 4). The driver can return the steering angle to the neutral position by lowering his or her toes. The drift error is canceled by performing this operation while the car is running straight.

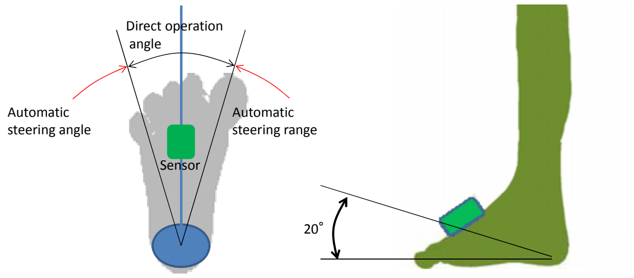


Figure 4. Example of controlling the steering by foot

E. Sensor attachment and actions

We measured the car control characteristics for several actions: rolling the ankle, moving the forefinger, moving the wrist, rolling the lower arm, moving the lower arm backward and forward, and moving the upper arm backward and forward. The position of the sensor and the motion of each body part are as follows.

[Rolling the ankle]

We considered using knee turning and knee movement for moving the gyro sensor. However, these movements produce a narrow movable angle, so we used rolling the ankle. The sensor is placed on top of the feet, as shown in Fig. 4. The sensor moves when the foot is pivoted right or left on the heel.

[Moving the forefinger]

The sensor is placed on the top portion of the top joint of the forefinger and is moved as shown in Fig. 5.

[Moving the wrist]

The sensor is placed on the back of the hand and is moved as shown in Fig. 6.

[Rolling the lower arm]

The sensor is placed on the lower arm and is rolled as shown in Fig. 7.

[Moving the lower arm backward and forward]

The sensor is placed on the back of the lower arm and is moved as shown in Fig. 7.

[Moving the upper arm backward and forward]

The sensor is placed on the upper arm and is moved as shown in Fig. 8.

 ISBN: 978-1-61208-250-9

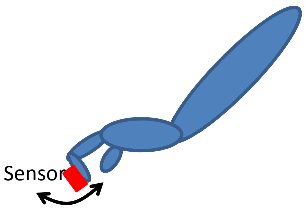


Figure 5. Moving the forefinger

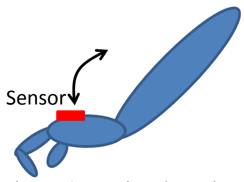


Figure 6. Moving the wrist

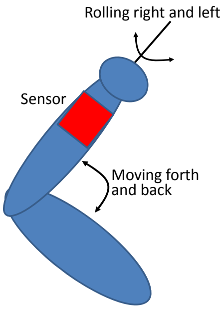


Figure 7. Rolling and moving the lower arm

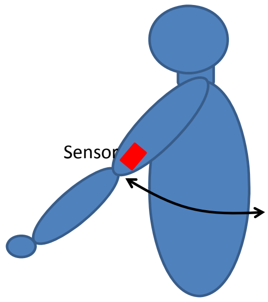


Figure 8. Moving the upper arm

IV. DRIVING SIMULATOR

We evaluated our driving interface by using a driving simulator. We used OpenGL [9] as the 3D program interface and developed the program using the "glut", "sdl" [10], "glew", and "OpenAL" tools [11]. The simulated car was operated by using a gyro sensor as the steering controller, the brake pedal, and the acceleration pedal.

A. Driving course

Our ultimate aim is to help disabled people obtain a driver's license, so we designed the driving course on the basis of a typical driving school course (Fig. 9). It comprised a rectangular outer course, two crank-shaped courses, two S-shaped courses, and two parallel parking courses. The outer course was 300 × 120 m and had a corner radius of 20 m. A 3.3-m-wide driving lane ran in each direction.

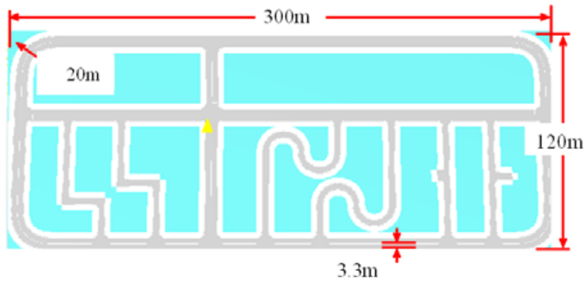


Figure 9. Driving course for simulation

B. Motions

Two motions were simulated: gyration and acceleration [13].

1) Gyration

A steady gyrating motion was applied to the car under the following assumptions.

- The movement of the car was the broadside motion of a rigid body. That is, the car was rigid and free of distortion.
- The speed was constant throughout each curve.
- The characteristics of the right-side tires were the same as those of the left-side tires.

The radius of the gyrating movement is given by the following equation in which V is the running speed and δ is the steering angle.

$$R = (1 + CV^2)\frac{1}{\delta} \tag{1}$$

The C is given by the following equation, in which the mass of the car is m, the cornering force on the front tires is K_f, that on the rear tires is K_r, the wheel base is l, and the distances between the car's center of gravity and the front and rear axles are l_f and l_r.

$$C = -\frac{m}{2l^2}\frac{l_f K_f - l_r K_r}{K_f K_r} \tag{2}$$

Each parameter was set to produce driving characteristics similar to those of an actual car. The cornering force was controlled by adjusting the radius of the gyrating movement: the larger the radius, the stronger the cornering force.

2) Acceleration

The acceleration A_c of an actual car depends on the engine torque, the transmission gear ratio, the tire radius, the vehicle weight, and the engine speed. The engine speed depends on the degree to which the accelerator pedal is pressed.

Air resistance R_a and rolling resistance R_r are considered to be the total running resistance.

$$R_a = \frac{1}{2}C_d \rho S V^2 \tag{3}$$

 ISBN: 978-1-61208-250-9

where C_d is the aerodynamic coefficient, ρ is the fluid density of the air, and S is the gross surface area of the car.

$$R_r = C_{rr} mg \quad (4)$$

where m is the mass of the car, C_{rr} is the rolling coefficient, and g is the gravitational acceleration. Resultant acceleration A is given by

$$A = A_c - (R_a + R_r) \quad (5)$$

C. Simulation display

An example view through the windshield is shown in Fig. 10. The upper right portion shows the position of the car on the course. The operation monitoring tool we developed to facilitate operation is shown in Fig. 11. It helps the driver recognize the angle of the sensor as the angle of the steering wheel and the angle of the toes. It also displays the degree to which the accelerator or brake pedal has been pushed.

Figure 10. Example view through front window

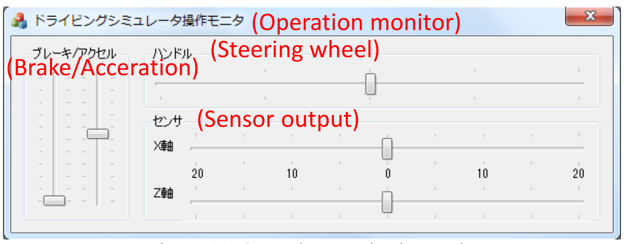


Figure 11. Operation monitoring tool

D. Measured data

Nine data items were measured.

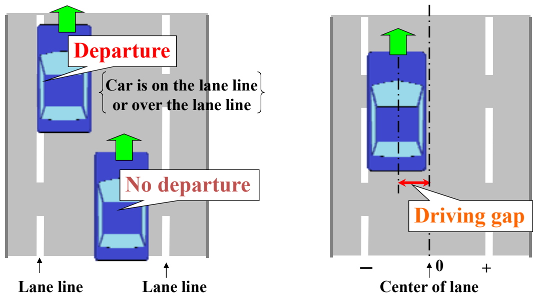
1) *Steering angle*
2) *Running speed*
3) *Distance driven and driving time*
4) *Position of car on course*
5) *Distance between left side of car and left-side lane marker line*
6) *Distance between right side of car and right-side lane marker line*
7) *Degree to which accelerator pedal was pushed*
8) *Depth to which brake pedal was pushed*
9) *Angle of car relative to driving direction*

V. Evaluation

A. Steering control

To analyze performance against the 2nd and 3rd steering operation requirements described in Section III-B, we calculated the ratio of lane departure (RLD) and the standard deviation of the driving gap (SDDG).

- As illustrated in Fig. 12(a), lane departure means that one or more of the tires run on or across a lane marker line. RLD is the ratio between the distance driven and the distance during which the car departed the lane.
- As illustrated in Fig. 12(b), the driving gap is the distance between the lane center and the car's center line. A value of zero means that the car is centered in the lane. SDDG was calculated using the values obtained for the car running on a straight portion of the course.



(a) Lane departure (b) Driving gap

Figure 12. Lane departure and driving gap

Since the first objective of this research is to develop a driving interface for arm and finger disabled people, we focused on foot-controlled steering, as illustrated in Fig. 4.

A non-linear function ($y = x^n$) was used to control the steering within the direct operation angle. We measured the position of car on course and calculated the RLD and SDDG for n=1 – 4 in $y = x^n$. We also measured and calculated the same data using a steering-wheel-type game controller for comparison. The measured and calculated data used for the non-linear function are listed in Table I for one of the four participants, a person with much experience driving a car using his foot with the proposed driving interface. Since the details of the experiment by Wada and Kameda were not published, we did not compare the performance of our control function with theirs.

We initially thought that a driver could easily operate the car by using semi-automatic steering control. However, the RLD and SDDG were much worse than with the game controller when the direct operation angle was ±15° and the corresponding steering wheel angle was ±30°. We observed that it was very difficult to drive the car using semi-automatic steering control during typical driving maneuvers. As illustrated in Fig. 13, neither of the two participants in this experiment negotiated the corners of the rectangular outer course smoothly under those conditions. We concluded that semi-automatic steering control is suitable only for parking and traversing a crank-shaped course and that non-linear steering control is better for typical driving maneuvers. As shown in Table I, RLD and SDDG were the smallest for n = 1. But, the difference in values between n = 1 and n = 3 is negligible. RLD was the smallest for n = 3 during the latter laps, as shown in Fig. 14. This means that a driver tends to continue to drive longer when n = 3.

We measured and calculated the same data for three other participants (students) for $y = x^3$ (DOA=±20°, SWA±180°) for comparison with $y = x^1$ (DOA=±15°, SWA±30°) to see whether other drivers showed the same tendencies as participant Y. There were four participants in total, but not all of them participated in each experiment. The measured and calculated data are listed in Table II. The values for "Other" are the averages for the other participants.

TABLE I. MEASURED AND CALCULATED DATA FOR NON-LINEAR CONTROL FUNCTION

	$y=x$	$y=x^2$	$y=x^3$	$y=x^4$	$y=x$	Game str. wheel
Operating body part	Left feet					
DOA*	±20°				±15°	
SWA**	±180°				±30°	
Distance driven (m)	791.3	790.9	790.9	790.9	795.4	790.0
Ave. speed (km/h)	26.6	26.3	25.5	27.3	14.3	30.4
RLD (%)	0.15	0.38	0.24	0.91	9.8	0
SDDG (m)	0.21	0.24	0.27	0.29	0.15	0.11

*DOA: Direct operation angle
**SWA: Corresponding steering wheel angle

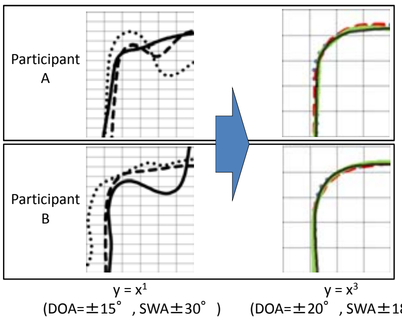


Figure 13. Cornering performance for participants A and B

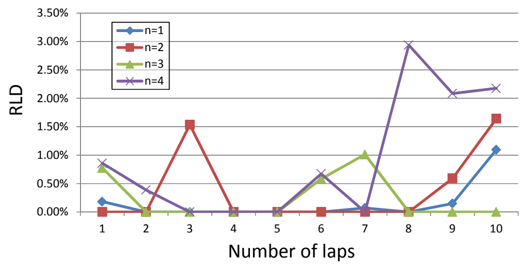


Figure 14. RLD against no. of laps

TABLE II. MEASURED AND CALCULATED DATA FOR MULTIPLE PARTICIPANTS

	$y = x^1$		$y = x^3$		Game str. wheel	
Participant	Y	Other	Y	Other	Y	Other
Operating body part	Left foot				Hand	
DOA	±15°		±20°		—	
SWA	±30°		±180°		—	
Distance driven (m)	795.4	807.0	794.3	791.9	790.0	794.5
Ave. speed (km/h)	14.3	18.3	28.1	27.8	30.4	33.5
RLD (%)	9.8	52.3	3.3	12.5	0	2.3
SDDG (m)	0.15	1.31	0.35	0.46	0.11	0.21

Each participant's RLD and SDDG for $y = x^3$ (DOA=±20°, SWA±180°) was better than for $y = x^1$ (DOA=±15°, SWA±30°). Example cornering performances for two of the participants are shown in Fig. 13. Although every participant smoothly traversed the corners when $y = x^3$ (DOA=±20°, SWA±180°), none of them achieved driving operation comparable to that achieved with a steering wheel. However, it should be possible to eliminate the difference and achieve operation closer to that with a steering wheel through more practice and experience.

We calculated the RLD for the crank- and S-shaped courses for four participants as well. The measured and calculated data are listed in Table III. The values are the averages for them. Traversing the crank- and S-shaped course was more difficult than traversing the rectangular course. The driver had to use both non-linear and semi-automatic steering control. Driving performance for these two courses differed greatly. As shown in Figs. 15 and 16, the performance of participant D was very close to that with the game steering wheel while that of participant C substantially diverged from it. This indicates that performance with the proposed steering operation interface should approach that with a steering wheel as the amount of practice and experience increases.

TABLE III. MEASURED AND CALCULATED DATA FOR CRANK- AND S-SHAPED COURSE

	Crank-shaped course		S-shaped course	
Operation device	Game wheel	Sensor	Game wheel	Sensor
DOA	±20°			
SWA	±180°			
Cont. function	$y = n^3$ and semi-automatic steering control			
Distance driven (m)	79.9	78.2	123.6	125.8
Ave. speed (km/h)	9.7	8.3	15.0	12.9
RLD-Ave. (%)	17.2	22.9	8.9	16.8
RLD-Max. (%)	26.6	42.8	21.2	40.8
RLD-Min. (%)	6.6	8.5	0	1.2

 ISBN: 978-1-61208-250-9

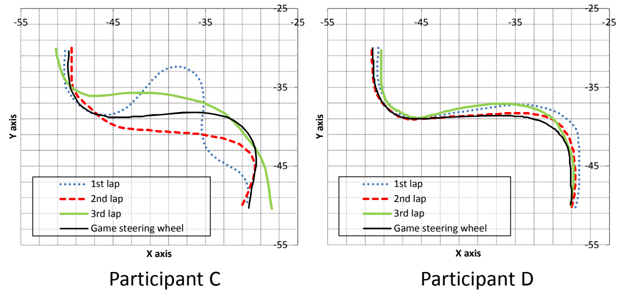


Figure 15. Crank-shaped course performance for participants C and D

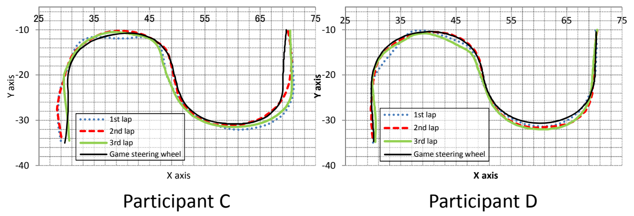


Figure 16. S-shaped course performance for participants C and D

B. Results by body part

To examine the applicability of the proposed driving interface to various types of disabilities, we measured and calculated the RLD and SDDG for steering control using the forefinger, the wrist, the lower arm, the upper arm, and the foot. The data measured for participant Y are listed in Table IV. The movements are illustrated in Figs. 5–8. The RLDs and the SDDGs for each body part were roughly the same except for the upper arm. The data for the upper arm was the worst because it is difficult to precisely move the upper arm (and shoulder). Although the forefinger had the largest DOA, its RLD was worse than that for the foot. The reason is that it is too easy to move the finger. An adapter might be useful in achieving steady forefinger movement.

TABLE IV. MEASURED AND CALCULATED DATA BY BODY PART

	Fore-finger	Wrist	Lower arm 1*	Lower arm 2*	Upper arm	Foot
DOA	±70°	±50°	±40°	±40°	±40°	±20°
SWA	±180°					
Cont. function	$y = n^3$					
Distance driven (m)	791.0	791.3	791.3	792.0	796.0	794.3
Ave. speed (km/h)	26.3	26.0	26.0	28.4	24.8	28.1
RLD (%)	3.8	3.2	4.7	4.2	17.6	3.3
SDDG (m)	0.35	0.26	0.26	0.26	0.33	0.35

*Lower arm 1: Moving lower arm forward and backward
*Lower arm 2: Rolling lower arm right and left.

VI. CONCLUSION

Our proposed steering operation interface for disabled people uses gesture operation. It incorporates both non-linear and semi-automatic steering control. Simulation experiments using foot control and a gyro sensor showed that semi-automatic steering control is suitable only for parking and traversing a crank- or s-shaped course and that non-linear steering control is better for typical driving maneuvers. The driving operation for each body part except for the upper arm was close to that achieved with a steering wheel. More practice in using the proposed interface should enable a user to achieve steering control closer to that with a steering wheel.

We plan to develop a prototype control device using a strain gauge instead of a gyro sensor and evaluate its driving operation for many participants. We also plan to evaluate our proposed interface in an actual car.

REFERENCES

[1] Development of Honda's Franz System Car; http://world.honda.com/history/challenge/1982franzsystemcar/index.html, January 2013.
[2] Joystick Driving System: allows wheelchair users to drive a car; http://www.itechdiary.com/joystick-driving-system-allows-wheelchair-users-to-drive-a-car.html, January 2013.
[3] Masayoshi Wada, Fujio Kameda, "A joystick car drive system with seating in a wheelchair", IEEE IECON '09, pp. 2163-2168, November 2009.
[4] FORUM8 Air Driving and RoboCar http://www.youtube.com/watch?v=LMr2dyfAzl0, January 2013.
[5] A.S.M. Mahfujur Rahman, Jamal Saboune, Abdulmotaleb El Saddik, "Motion-path based in car gesture control of the multimedia devices," ACM DIVANet '11, Proceedings of the first ACM international symposium on Design and analysis of intelligent vehicular networks and applications, pp.69-75, November 2011.
[6] Tanja Döring, Dagmar Kern, Paul Marshall, Max Pfeiffer, Johannes Schöning, Volker Gruhn, Albrecht Schmidt, "Gestural interaction on the steering wheel: reducing the visual demand," ACM CHI '11, Proceedings of the 2011 annual conference on Human factors in computing systems, pp.483-492, May 2011.
[7] Yoshitoshi Murata, Nobuyoshi Sato, Tsuyoshi Takayama, Shinetsu Onodera, "A Gesture-based Remote Control for Finger Disabled People", IEEE, GCCE 2012, pp.411-415, October 2012.
[8] Samsung SMART-TV, http://www.samsung.com/us/2012-smart-tv/, January 2013.
[9] OpenGL – The Industry Standard for High Performance Graphics, http://www.opengl.org/, January 2013.
[10] Simple DirectMedia Layer, http://www.libsdl.org/, January 2013.
[11] Home – OpenAL, http://connect.creativelabs.com/openal/default.aspx, January 2013
[12] Masato Abe, "Automotive Vehicle Dynamics - Theory and Applications," Tokyo Denki University Press, (2008).
[13] Giancalro Genta, "We apply the steady gyrating movement to a car under following assumptions," World Scientific Publishing, 1997

The European MobileSage Project – Situated Adaptive Guidance for the Mobile Elderly

Overview, status, and preliminary results

Till Halbach Røssvoll
Norwegian Computing Center (Oslo, Norway)
till.halbach.rossvoll@nr.no

***Abstract*—MobileSage is an AAL project with the goal to provide particularly elderly people with relevant and useful multimodal guidance on demand, depending on the context, and in a personalized, adaptive, accessible, and usable manner. The project aims to increase the independence of elderly people in particular in the home environment and during travel. This work presents the project and its two main deliverables, discusses related research, summarizes preliminary results, and gives an outlook of anticipated results.**

Keywords—AAL, mobile, application, assistance, help on demand, personalization, adaptive, accessible, usable, multimodal

I. Introduction

Today's society is flooded with a plethora of new technological devices. Although easy to use for newer generations, they are often unfamiliar to and can pose severe difficulties for the elderly. This is why many senior citizens meet these ICT based solutions (Information and Communication Technology), such as automated ticketing machines, with anxiety.

At the same time, modern elderly live longer, they are healthier, more active, mobile, independent, and more demanding customers than ever before [1]. They are increasingly looking for useful, user friendly and personalized ICT services that add value to their active and mobile life, and that can help them to stay active despite various impairments. Here, MobileSage provides a timely approach and solution. The project was started in July 2011 and ends by January 2014.

This article provides an update for a previous work-in-progress paper [2] and is organized as follows: After the European AAL Joint Programme has been introduced, related projects and research is discussed, before the MobileSage Project is presented together with its two main service deliverables. After that, preliminary results are discussed, an overview of the way ahead is given, and finally the conclusion is drawn.

II. The European AAL Programme

The European Ambient Assisted Living Joint Programme (AAL JP) provides the settings for this ICT based innovation project targeting elderly individuals [3]. The main goal of AAL is to improve "*the quality of life, autonomy, participation in social life, skills, and employability of older people*", while service delivery enhancement and care costs reduction are secondary targets. MobileSage targets the program's third call, named *ICT-based Solutions for Advancement of Older Persons' Independence and Participation in the "Self-Serve Society"* [4]. The Call considers both *primary* end-users (elderly individuals), *secondary* users (family members, care givers), and *tertiary* end-users (NGOs, public authorities, companies, vendors).

III. Related Work

Before the development in MobileSage started, a literature study was carried out to get an update about state of the art in the field. The study revealed that there are few, if any, existing solutions witch the same scope as the project. Especially rare are related projects and solutions of this particular kind directed towards elderly and impaired users. However, the study showed that there are a number of recent and ongoing international projects related in various ways to the topics addressed by MobileSage; they are discussed subsequently.

A number of international projects fall within the same *scope*. The APSIS4ALL Project is dealing with personalizing public digital terminals such as ATMs and ticket machines [5], where an adaptive interface and personalized interaction is achieved by the human communicating with the terminal through a smartphone. In the ASK-IT Project, a framework has been developed that employs personalized intelligent agents for service provision and search for suitable semantics [6]. The MyUI Project is addressing specific user needs towards ICT products in general through adaptive personalized interfaces, and by recording user behavior and context information in real-time [7]. GUIDE is a project to design tools and aids for developers to efficiently integrate personalization, user friendly interaction, and accessibility features into applications [8]. In the DIADEM Project, electronic/online forms were made adaptable to the cognitive skills of the user by monitoring user actions and tailoring the user interface based on these data [9]. The GPII Project is building a framework that allows to store universal user profiles in the cloud [10]. Later, the profile can be accessed to adapt the user interface of any device to a user's needs and preferences.

The following research projects are related to MobileSage with regard to *user interface* matters. The aim of SNAPI was to develop a data format for the storage of user profiles, with a focus on smartcards and public digital terminals [11]. In the GoldUI Project, elderly are offered a num-

 ISBN: 978-1-61208-250-9

ber of cloud services which are deemed useful in the everyday life [12]. The goal then is that these services can be accessed through a variety of platforms, including telephone, smartphone, tablet PC, TV, and radio.

The following research projects consider *transportation.* WayFiS is addressing the topic of travel challenges [13], in terms of a web based pre-planning service and a mobile application, with route calculations, taking into account an individual's desired physical activity, nutrition needs, necessary facilities along the route, and disease restrictions, while trying to avoid inaccessible places. In Mediate, a number of criteria and tools have been developed to measure accessibility in public transport, including accessible ticketing and information systems, and accessibility information of public transport system in Europe [14]. In the Access2all Project, a number of guidelines, recommendations, roadmaps, and new research initiatives have been worked on [15]. The AmbienNet Project has developed an indoor location system based on intelligent infrastructure and a sensor network [16].

The following projects concern *multimodality.* In the HAPTIMAP Project, a cross-platform toolkit for the design of user interfaces incorporating haptics, audio, and video input and output is being developed for retrieval, storing, and manipulation of geographic data [17]. In the course of the MAPPED Project, a mobile application has been developed that provides accessibility information on buildings, traveling means like buses, trains, etc., combined with localization techniques [18]. Finally, the HearMeFeelMe Project [19] was aiming at replacing visual and textual information with audio, combined with touch-based interfaces for information access, all integrated in a mobile application employing near-field technology to gather information about items to buy, such as food and medication [20].

IV. Overview of MobileSage

The idea of MobileSage is to provide elderly with a personalized context-sensitive tool which provides relevant, accessible, usable, and multimodal assistance for carrying out and solving everyday tasks and problems in the self-serve society whenever and wherever they occur, just in time, or, on demand [21], [22]. As such, the system acts like a facilitator of knowledge. The project addresses all the areas of interest of the 3rd AAL Call, and it also accounts for the three aforementioned user groups. MobileSage consists basically of two services, which are the main two deliverables of the project. They are called Help-on-Demand (HoD) Service and Content Management (CM) Service. Figure 1 shows the overall architecture and the major building blocks.

The content itself is fetched from the CM Service. There are no limitations regarding what kind of content can be facilitated, including manuals, usage instructions, and travel descriptions. It is anticipated that mainly secondary users will upload content to the CM Service as they are likely to have a direct interest in helping the primary users. However, it is expected that also tertiary users provide content. For instance, a manufacture might regard it as advantageous to provide manuals for their ticket machines, or the railway operator that runs these machines might do so. Even a municipality might be interested in producing such help content as a special service for their citizens. Of course, there is nothing that prevents literate primary users from producing and making help content available themselves through the CM Service.

A. Help-on-Demand Service

The Help-on-Demand (HoD) Service is a thick-client application, a personal agent, running on a smartphone. Its main building blocks are built up in a service oriented manner, see Figure 1. The user interacts with the Dialog Manager through the User Interface. The Dialog Manager utilizes the functionality provided by the Profile Service, which takes care of the user profile. Besides the user's personal preferences also the use pattern is stored there. The behavior of the user and actions in the User Interface are logged and analyzed by the Personalization Service which is responsible for adjusting the user profile accordingly.

Besides utilizing services provided as cloud services, the Dialog Manager is further in contact with the Reasoning Service with the task to make an educated guess of the user's context. In order to do so, the Reasoning Service itself makes use of network services such as Media Service, Search Service, Content Service, etc. The most important factor for determining the user context is the current location. The Reasoning Service therefore gets help from the Localization Service which has the responsibility to determine the user's location with highest possible accuracy, based on technologies like A-GPS, WLAN, GSM/GPRS, and NFC, and triangulation methods.

B. Content Management Service

The Content Management (CM) Service is organized as a cloud service and runs on a web server. A user who wants to produce multimodal content uses a User Agent like a browser to gain access to the service. The user interacts with the service's Dialog Manager, which in turn controls the User Interface. The main logic for handling the multimodal content lies in the Content Manager, which has a modular design to be able to add additional modalities in a simple way. MobileSage includes the modules Video, Audio/Speech, and Text, including Subtitles. The fabricated content is finally stored by the Content Service. It is also possible, though, to refer to content which is located elsewhere.

V. Current state and preliminary results

Beta versions of the smartphone application and the web service have recently been released for evaluation purposes to complete the first development iteration. Currently, field trials with relevant user groups are being carried out in the three countries Spain, Romania, and Norway. The trials comprise mostly qualitative studies where the user has to solve a particular task, such as operating a coffee machine.

Another scenario involves a travel situation: A user has to find a ticket machine at a subway station, pay for a ticket, and find the right subway platform. Appropriate content is found by means of NFC tags or QR codes placed out in the field, or by a user issued search with a particular search phrase. Both methods will initiate a geographically limited database search for relevant content in the user's preferred languages and modalities. Available results are presented in the smartphone app for the user who can then in turn pick the one he/she thinks suits the context best, be it text, formatted text, audio, video, captioned video, images, etc.

 ISBN: 978-1-61208-250-9

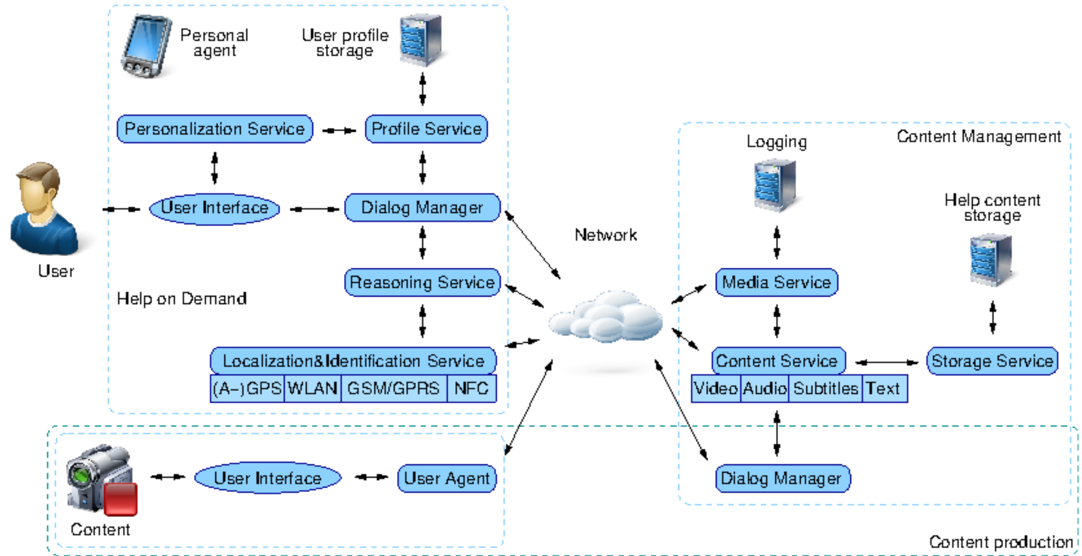


Figure 1: System architecture for Help on Demand Service (left) and Content Management Service (right) with the network in between

Results from these trials are going to be published in 2013. Concerning other publications, the following project results have been published so far:

A. Deliverable D2.1: User Needs Analysis

At the start of the project, user focus group work was conducted in the three countries Norway, Romania, and Spain to cover the needs of primary users [23]. The in total 39 participants of the focus groups were selected to represent a broad range of parameters, including age (48 to 96), gender (24 female vs. 15 male), dis-/abilities (sensory and cognitive impairments), nationality (4 foreigners), and ICT experience and usage. The participants were given the description of 2 scenarios. One scenario dealt with a travel situation, where a user was traveling in a foreign country, encountered language problem, and had to cope with reduced vision. The other scenario comprised a challenge occurring in the home environment, where an elderly user faced the problem of understanding the technical manual of an electric household appliance.

The results from the focus group work show that the target group "elderly persons" indeed is a very heterogeneous group with a wide range of – partly opposite – needs and wishes. This applies also to the users' familiarity with ICT in general and mobile technology in particular, which ranges from none to professional users.

An in-debth discussion of the results is available in the respective deliverable [23].

B. Deliverable D2.2: User Requirements Specification

The results from the user needs analysis were collected and formulated as user requirements [24]. The roughly 50 requirements address the expectations of primary users towards HoD and CM. An example of a user requirement is "*The HoD Service shall have measures for privacy protected audio output*". The user requirements served as input to the process of formulating the first draft of the system requirements for the two main deliverables of MobileSage.

C. Deliverable 2.3: System Requirements Specification for the Help-on-Demand System

This specification addresses the requirements towards the HoD application regarding the user interface, system functionality, and input/output matters. Its informal parts also include user interface mockups [25]. An example of a system requirement is "*An emergency button shall be shown on any screen of the application if the value for this feature has been set to 'on' in the user profile*". This deliverable serves as a set of instructions for the system development and integration.

D. Deliverable 2.4: System Requirements Specification for the Content Management System

This specification addresses the requirements towards the CM Service regarding the user interface to the database, system functionality, and input/output matters, including the data exchange with HoD and user interface mockups [26].

E. HoD and CM software prototypes

As already mentioned, beta versions of the smartphone application and the web service have been released.

F. Deliverable D 5.3: Market Analysis & Socio-Economic Impact and Potential

This document gives an in-debth analysis of the market segments that are relevant for the HoD application, stakeholders, and possible business models [27]. It also discusses similar products which already are on the market, and related projects.

VI. OUTLOOK

The development process in MobileSage foresees multiple iterations for the HoD and CM services, where one iteration consists of the steps specification work, risk analysis, development and integration, and evaluation. The development is user centric, meaning that users from the target groups are involved in each evaluation step, and that the development starts with an analysis of the user needs. The idea is that the user requirements, which are a direct consequence of the user needs analysis, are “translated” to system requirements, and that the specification of those is refined in each development iteration based on the results from the previous user evaluation. Thus, deliverables D2.3 and D2.4 are expected to undergo several new revisions, as will the software prototypes. For an updated schedule and timeline, please refer to the project's web site [21].

VII. Conclusion

In this article, the AAL project MobileSage has been introduced by presenting its two main service deliverables, as well as status and preliminary results, and an outlook on the road ahead. The work also briefly mentioned the AAL Programme and discussed related projects and research.

The objective of MobileSage is to increase digital participation and involvement by providing particularly elderly people with relevant and useful multimodal guidance on demand, depending on the context, and in a personalized, adaptive, accessible, and usable manner. As such, it addresses all of AAL's areas of interest. The project's overall goal is to significantly increase the independence of elderly people in the home environment and during travel. It aims at stimulating people to help themselves and easing the production and availability of help content for just about anything. A first beta version of the services has already been released and is undergoing testing in Q4 2012, while the project itself is planned to last throughout 2013.

Acknowledgment

This work is partly funded by the European Commission through the AAL Programme, the Norwegian Research Council, and National bodies in Spain and Romania. The author would like to thank the consortium members for their valuable contributions and all individuals involved in the user studies for their feedback.

References

[1] Metz D. & Underwood M., *Older Richer Fitter. Identifying customer needs of Britain's ageing population*. Age Concern (2005)
[2] Røssvoll T.H.. *The European MobileSage Project --– Situated Adaptive Guidance for the Mobile Elderly*. In Electronic Government and Electronic Participation, Joint Proceedings of Ongoing Research and Projects of IFIP EGOV and IFIP ePart 2012 (2012)
[3] AAL Association. *Innovative ICT solutions for Ageing -- Ambient Assisted Living*. http://www.aal-europe.eu/ (Last accessed: 2012-06-30)
[4] AAL Association. *ICT-based Solutions for Advancement of Older Persons' Independence and Participation in the "Self-Serve Society"*. 3rd Call for Proposals (2010)
[5] APSIS4ALL Project Consortium. *The APSIS4ALL Project*. http://www.apsis4all.eu/ (Last accessed: 2012-06-30)
[6] ASK-IT Project Consortium. *The ASK-IT Project*. http://www.ask-it.org (Last accessed: 2012-06-30)
[7] MyUI Project Consortium. *The MyUI Project*. http://www.myui.eu (Last accessed: 2012-06-30)
[8] Guide Project Consortium. *The Guide Project*. http://www.guide-project.eu (Last accessed: 2012-06-30)
[9] DIADEM Project Consortium. *The DIADEM Project*. http://www.project-diadem.eu (Last accessed: 2012-06-30)
[10] GPII Project Consortium. *The GPII Project*. http://gpii.org (Last accessed: 2012-06-30)
[11] Snapi Project Consortium. *The Snapi Project*. http://www.snapi.org.uk (Last accessed: 2012-06-30)
[12] GoldUI Project Consortium. *The GoldUI Project*. http://www.goldui.eu (Last accessed: 2012-06-30)
[13] WayFiS Project Consortium. *The WayFiS Project*. http://www.wayfis.eu (Last accessed: 2012-06-30)
[14] Mediate Project Consortium. *The Mediate Project*. http://www.mediate-project.eu (Last accessed: 2012-06-30)
[15] Access2all Project Consortium. *The Access2all Project*. http://www.access-to-all.eu (Last accessed: 2012-06-30)
[16] Abascal J., Sevillano J.L. & Artigas J.I., *AmbienNet - Ambient Intelligence Supporting Navigation for People with Disabilities*. Jornada de Seguimiento de Proyectos (2009)
[17] HAPTIMAP Project Consortium. *The HAPTIMAP Project*. http://www.haptimap.org (Last accessed: 2012-06-30)
[18] MAPPED Project Consortium. *The MAPPED Project*. http://services.txt.it/MAPPED (Last accessed: 2012-06-30)
[19] Ervasti M., Isomursu M. & Leibar I.. *Touch- and Audio-based Medication Management Service Concept for Vision Impaired Older People*. In IEEE International Conference on RFID-Technologies and Applications (RFID-TA) (2011)
[20] Vastianos G., Kyriazanos D., Segou O., Mitilineos S. & Thomopoulos S.. *Indoor Localization Using Passive RFID*. In Proceedings of Signal Processing, Sensor Fusion, and Target Recognition (2011)
[21] Sánchez V.. *The MobileSage Project*. http://mobilesage.eu (Last accessed: 2012-06-30)
[22] Solheim I. & Røssvoll T.H.. *MobileSage Project Factsheet*. http://nr.no (Last accessed: 2012-06-30)
[23] Dale Ø.. *MobileSage Deliverable D2.1: User Needs Analysis* (2012), http://mobilesage.eu/public-documents/public-deliverables
[24] Røssvoll T.H.. *MobileSage Deliverable D2.2: User Requirements Specification* (2012), http://mobilesage.eu/public-documents/public-deliverables
[25] Curescu L., Anghelache I. & Røssvoll T.H.. *MobileSage Deliverable D2.3: System Requirements Specification for Help-on-Demand Service* (2012), http://mobilesage.eu/public-documents/public-deliverables
[26] Røssvoll T.H. & Gracia V.A.. *MobileSage Deliverable D2.4: System Requirements Specification for Content Management Service* (2012), http://mobilesage.eu/public-documents/public-deliverables
[27] Sánchez V. & Sánchez J.. *MobileSage Deliverable D 5.3: Market Analysis & Socio-Economic Impact and Potential* (2012), http://mobilesage.eu

 ISBN: 978-1-61208-250-9

e-Learning Environment with Multimodal Interaction

A proposal to improve the usability, accessibility and learnability of e-learning environments

André Constantino da Silva
Institute of Computing (PG)
UNICAMP, IFSP
Campinas, Brazil, Hortolândia, Brazil
acsilva@ic.unicamp.br

Heloísa Vieira da Rocha
Institute of Computing, NIED
UNICAMP
Campinas, Brazil
heloisa@ic.unicamp.br

***Abstract*—The Human-Computer Interaction is challenging the use of many modalities to interact with an application. The e-Learning environments interfaces are been exposed to this diversity of modalities, but they are designed to be used with a limited set. The impact is that users have interaction problems caused by the cross modality. The e-Learning environments need to evolve allowing users to interact with a more broadly interaction styles. One solution is adopt Multimodal Interaction concepts, improving the usability and accessibility of the e-Learning environment and make possible to embrace better learning contexts, property that we define as learnability.**

Keywords-Human-Computer Interaction; Interaction Styles; Multimodal Interaction; Electronic Learning Environment.

I. INTRODUCTION

The Human-Computer Interaction is challenging the replacement of the mouse and actual interfaces for interfaces that works with natural interaction, non-tactile interfaces, speech recognition, facial and movement recognition and gestures [1][2]. So, there are many ways to interact with digital artifacts and applications, like keyboards, mouse, small, medium and big displays, voice, touch and gestures. Many hardware components are available supporting different interaction styles.

The Multimodal Interaction is a solution to possibility the use of an application in this diversity of interaction styles, allowing users interact with computers by many input modalities (e.g., speech, gesture, eye tracking) and output channels (e.g., text, graphics, audio) [3]. Multimodal Interaction is proposed to turn the human-computer interaction more natural, i.e., more close to the human-human interaction. The main benefices are the increase of application's usability, accessibility, flexibility and convenience [3]. But, building a multimodal interaction system is not a trivial task yet, because the literature does not have sufficient information about how to design this kind of system and there is a lack of technologies to support them.

e-Learning environments like Moodle [4], SAKAI [5], TelEduc [6], Ae [7] are applications that use the Web infra-structure to support teaching and learning activities. The e-Learning environments are designed to support a variety of users and learning contexts, but they are designed to support a limited interaction styles, usually keyboard and mouse as input and a medium screen as output.

To attend this demand, the e-Learning environment needs to have good usability, accessibility, mobility and learnability. Considering all these dimensions is not a trivial task. Does the multimodality can improve these requirements on e-Learning environments?

Section II presents a literature review about e-Learning environments, multimodality and multimodal interfaces. Section III presents the research problem that we want to deal, and Section IV our hypothesis and methods. Section V some preliminary results and expected contributions.

II. LITERATURE REVIEW

The actual versions of e-Learning environments take advantages of the Web to offer content with text, images, audios and videos in a hypertext document. Tools like chat, forums, portfolios, repositories are widely used, and tools those explore the audio and video resource to user communication, such as instant messenger and video-conferences, are becoming common among the environments.

Due the diversity of users whom may use the e-Learning environments, these systems need to have good usability so that the user interface does not prejudice the teaching and learning activities. Accessibility is another important requirement to allow disabled people to use the environment. So usability and accessibility are two desired requirements to the e-Learning environments.

Devices, such as smartphones and tablets, are becoming increasingly popular; most of them have touch screen displays, access to the Internet and enough computing power to process Web pages. So, Web sites and Web applications, initially developed to be used with keyboard, mouse and a medium size display, are been accessed by small touchscreen devices. This is another aspect of accessibility, so the environments´ development teams are building solutions to provide access on mobile devices. Three kind of solution are emerging: specific device application; web site specific for mobile devices; and improve the web site for mobile and desktop access [8].

Two motivations allow the participants interacting anytime and anywhere with the content and each other; but, due to the device restrictions, there are needs to obtain better design solutions. The actual user interface design techniques take account just a limit set of input and output hardware, limited to the context, such as techniques to design user interface for desktop or for mobile platforms. But, there is a

 ISBN: 978-1-61208-250-9

lot of input or output hardware in these devices and these techniques are asked to consider all of them. Some input and output devices are: touchscreen, microphones, pen sensitive screen, touchpad, TrackPoint, accelerometers, joysticks, loudspeakers, small screen, large screen, printers, etc.

Another e-Learning environment characteristic is to be used in many of learning context, e.g., teacher training, undergraduate courses, and team training in all areas of knowledge. We call these as learnability. But, the actual hardware increases the difficulty to use the environment to produce content for any area and support student activities.

To attend this demand, the e-Learning environment needs to be usable and accessible for many users in many social, physical, technological and learning contexts. So, e-Learning environment needs to be evaluated in the usability, accessibility, mobility and learnability dimensions, a not trivial task.

Since the e-Learning environments were building to Web, they have a common architecture: the client-server. Client is responsible to render the user interface through a browser. It is in the client side that the user interacts with the system using input and output hardware. The server is responsible to process client´s requests and data persistence. The server knows few about the input and output devices in client side.

Ae is an e-Learning environment developed by a consortium of Brazilian universities using J2EE technology and component-based development process. Layered component-based software architecture was defined for Ae environment [9] with the following layers:

Presentation layer: provides the application user interface;

System layer: provides an interface for the application functionality, that is, it is a façade for the application business rules;

e-Learning layer: provides the component interfaces that implement the application's business rules, which can be used by various applications and which use services and functionalities from the infrastructure layer to implement the business rules;

Infrastructure layer: implements a set of infrastructure services such as, for example, data persistence;

Common services layer: has the public services that can be utilized and accessed by all other architecture layers, except the presentation layer.

Multimodal interaction is a research proposal to turn the interaction between humans and machines more natural, i.e., more close to the interactions between two humans, and have the benefits to increase the usability, flexibility and convenience [3].

Modality is the used term to define a mode what the user data input or a system output is expressed. The communication mode refers to the communication model used by two different entities to interact [10]. Nigay and Coutaz [11] define modality as an interaction method that an agent can use to reach a goal, and it can be described in general terms such "speech" or in specific terms such "using microphones".

For monomodal systems, designers are not limited to choose only one modality. But, in multimodal systems, they can choose many modalities, that used together, increasing the system flexibility and gives other benefices. Interfaces with this characteristic are called as multimodal interfaces and the system are called multimodal interaction systems. Mayes [12] defines multimodal interaction systems as a system with the capacity to communicate with the user by many different communication modes, using more than one modality, automatically gives or extracts mean.

According to Oviatt [10] "Multimodal interfaces process two or more combined user input modes (such as speech, pen, touch, manual gesture, gaze, and head and body movements) in a coordinated manner with multimedia system output".

Lalanne *et al.* [3] describe multimodal interaction systems, or multimodal systems, allow users to interact with computers though many data input modalities (e.g., speech, gesture, eye gaze) and output channels (e.g., text, graphics, sound, avatars, voice synthesis).

Bangalore and Johnston [13] say the critical advantage of multimodal interfaces is that they allow user input and system output to be expressed in the mode or modes to which they are best suited, given the task at hand, user preferences, and the physical and social environment of the interaction. Allowing users interact with many modes, it is possible to improve the accessibility because a multimodal interface can be used for a disabled person using the interaction mode that she can handle.

Multimodal content is common on multimedia systems. The research problem that we want proposes a solution is to use multimodality on user interaction and get benefits of multimodal content too.

Dumas, Lalanne and Oviatt [14] present a generic architecture for multimodal systems, turning more easy to understand the mainly components of the multimodal systems: a fusion engine, a fission module, a dialog manager and a context manager, which all together form what is called the "integration committee". The authors define "input modalities are received though various recognizers, which output their results to the fusion engine in charge of giving a common interpretation of the inputs. A fusion machine gives an interpretation for the data and it communicates it to the dialog manager, in charge of identifying the dialog state, the transition to perform, the action to communicate to an application, and/or the message to return through the fission machine. The fission machine returns a message to the user through the most adequate modality or combination of modalities, depending on the user profile and context of use. For this reason, the context manager, in charge of tracking the location, context and user profile, closely communicates any changes in the environment to the three other components, so that they can adapt their interpretations". Multimodal systems need to take account of all input done by the user to identify and process the solicited action.

Developing multimodal interaction systems is a complex task [14]; but, to turn more easy the development there are some frameworks, such as the ICARE framework [15], FAME [16] and special approaches [17][18].

Bouchet, Nigay and Ganille [15] define formally the CARE properties to characterize and assess aspects of multimodal interaction: the Complementarity, Assignment, Redundancy, and Equivalence that may occur between the interaction techniques available in a multimodal user interface. To aim build multimodal system, the authors propose the ICARE framework.

Larson [19] shows three general questions to response with a web application will be improved with a new mode of input: the new mode needs to add value to the web application, the application leverages the strengths of the new mode and avoids its weaknesses, and the users need to have the required hardware and software.

To implement multimodal system for web it is necessary consider both architecture: for multimodal systems and for web systems. Gruenstein, McGraw and Badr [20] present a framework to develop multimodal interfaces for web, the WAMI Toolkit. The framework defines tree client-side components (Core GUI, GUI Controller and Audio Controller) and more four server-side components (Web Server, Speech Recognizer, Speech Synthesizer, and Logger). The user interact with the Core GUI, described at HTML and JavaScript, and the Audio Controller, a Java Applet to receive the audio input. The collected data is sent to server to be treated by the Speech Recognizer and the Web Server components.

Zaguia *et al.* [21] present an approach to develop multimodal systems using fusion machines dispose on web services in such a way the user can choose the modalities that she sees fit to her situation instead of already pre-defined modalities from the beginning.

But we need to not only build a multimodal system; we are worry about the environment usability so that the interface does not prejudice the teaching and learning activities. Nielsen [22] defines usability as a combination of five elements: easy to learning, efficient, easy to remember, low probability of users do mistakes and user satisfaction.

III. THE RESEARCH PROBLEM

Since the number of devices accessing the Web grows, the e-Learning environments are exposed to a variety of interaction styles, including ones that are not considered in the design time. Just supporting these interaction styles causes cross modality interaction problems [8], limitation in the multimodality use, and do not take advantages from the interaction style in use. So it is necessary to develop a solution without these limitations.

Thinking about the learning domain, we ask "how do the users interact with a multimodal e-Learning environment? Is this a solution to allow an application be accessed by different devices with a diversity of users, physical and social contexts avoiding cross modality problems and get the better use of the interaction styles?"

Since there are no one multimodal interaction e-Learning environment to aim us to response these questions, we need built one. So we want to know how to develop a multimodal interaction e-Learning environment? Which are the cross modality problems related with this context? How can we identify?

It is the main problem, but several others derive: Is it possible to improve an application to be used by many interaction styles and get advantages? Which kind of modifications is necessary to get the best use of an interaction style? How to guaranties that the usability will not be affected? After the modifications, the application has a better accessibility? Allowing many interaction styles, do we have a new kind of application? How to distinguish the applications that have these characteristics from the other that does not have?

Due the multimodality allows users interact with many modes, maybe the user will use this mode not only to interact with the application, but to create content. How this impact in the environment architecture?

The special interest in the learning domain is the necessity to improve the e-Learning environment to better attend the teaching and learning activities in the variety of learning context, reaching out the maximum of learnability.

IV. THE RESEARCH HYPOTHESIS AND METHODS

We argue that it is possible improve the usability, accessibility and learnability of an e-Learning environment adding multimodality concepts in the environment´s user interface.

We planned to use empirical method to validate our hypothesis, building a multimodal interaction e-Learning environment prototype and doing user study, collecting interaction data and user opinions by observation and questionnaires. Not one interaction style will be studied per time; we want study various interaction styles being used to complete some tasks in the environment. The collected data will be used to verify the e-Learning environment in three dimensions: usability, accessibility and learnability.

Due to the quantity of interaction styles and the complexity to build a multimodal system, we need to define an incremental process to build the prototype, taking one or a limited number of interaction styles per time. But, it is important to reduce the development efforts for the next iteration, when another interaction style will be selected. Here, we will apply some software engineering techniques, like software components. We proposed these steps for this process:

1. Select an interaction style;
2. Do user studies to collect interaction problems using the selected interaction style and how the manner to realize the tasks change;
3. Redesign the software user interface to defining better solutions for identified problems;
4. Analyze the software to identify the components responsible to user interface and components that manipulate user data;
5. Find or implement recognizers and synthesizers for the selected interaction style;
6. Change the software architecture to have the main components of a generic multimodal system;
7. Do tests to collect errors and fix them.

We believe if we can do these steps two or three times; so, there is possible to repeat it until all interaction styles is considered.

To prove our solution, we propose to implement the Multimodal Interaction concepts in the Ae e-Learning environment [7], and use this new kind of environment to research advantages and disadvantages of multimodal interfaces in learning. We planned study the interaction using pen, touch and gesture in two tools of the Ae e-Learning environment: the Weblog and the Whiteboard. Both tools are used to construct the content and are selected based on our premises that they are good choices to study the multimodality.

V. PRELIMINARY RESULTS AND EXPECTED CONTRIBUTIONS

One of the preliminary results were some problems in the TelEduc environment that happened when users interact with touchscreen devices [8], i.e., some problems happened due the platform changing (when user access the environment using a smartphone) and some problems happened due the modality changing. The problems identified in TelEduc due the modality changing will happen on Ae environment.

Other contribution is related with how the multimodal concept changes the architecture of an application and how to find solutions. The Ae architecture needs to be changed to adopt the multimodal concepts. Considering the Web-Accessible Multimodal Interfaces architecture [20] and the architecture of multimodal systems [14], we redesign the Ae architecture (Fig. 1). Due to the fact that the browser has the responsibility to show the GUI components in the client side, some components are added to treat the data input from the input devices. The user interaction data is sent to the server, who have the responsibility to process this data and gives an interpretation for the received data. After the input data interpretation process, the correspondent action is sent to the system component. So, the fusion and fission machines will be on the server. The fusion machine will be called when the server receive the data from the client-side components, and the fission machine will be called when the system response the request, after the data processing. Since, we are using a component-based architecture, the presentation layer and the multimodal components can run in a proper server, increasing the scalability and performance.

The main expected contribution is to know more about the relationship between usability, accessibility and multimodality. For the educational context, we want to know if the support more interaction styles there is an impact in the learning contexts, so define the learnability.

The prototype is another contribution, because there is no one multimodal interaction e-Learning environment, we will call it |Ae|.

ACKNOWLEDGMENT

The authors thank the CAPES and CNPq for financial support and for FAPESP through TIDIA-Ae Project (n. 05/60572-1), which provided the equipment.

REFERENCES

[1] L. Kugler. “Goodbye, Computer Mouse”. Communications of ACM, vol. 51, n. 9, September 2008, pp. 56.

[2] B. Hayes. “Gartner´s Seven IT Grand Challenges”. Communications of ACM, vol 51, n. 7, July 2008, pp. 10.

[3] D. Lalanne, L. Nigay, P. Palanque, P. Robinson, J. Vanderdonckt, and J. Ladry. “Fusion Engine for Multimodal Input: A Survey”. Proc. of the 11th International Conference on Multimodal Interfaces (ICMI-MLMI´09), ACM Press, Nov. 2009, pp. 153-160, doi: 10.1145/1647314.1647343.

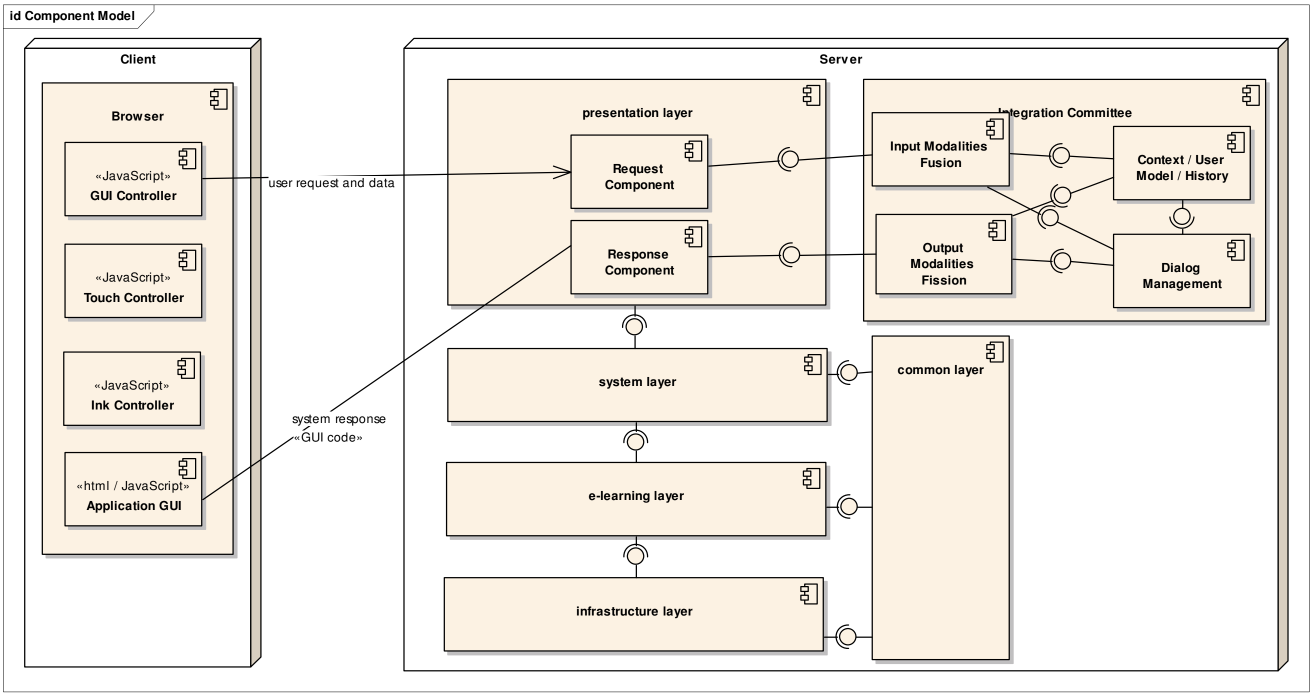


Figure 1. Ae architecture with multimodal components.

[4] Moodle Trust. “Moodle.org: open-source community-based tools for learning”. Available at <http://moodle.org>. [retrieved: Jan. 2013]

[5] SAKAI Environment. “Sakai Project | collaboration and learning - for educators by educators”. Available at <http://sakaiproject.org>. [retrieved: Jan. 2013]

[6] TelEduc Environment. “TelEduc Ensino à Distância”. Available at <http://www.teleduc.org.br>. [retrieved: Jan. 2013]

[7] Ae Project. “Ae - Aprendizado Eletrônico Environment”. Available at <http://tidia-ae.iv.org.br/>. [retrieved: Jan. 2013]

[8] A. C. da Silva, F. M. P. Freire, and H. V. da Rocha. “Identifying Cross-Platform and Cross-Modality Interaction Problems in e-Learning Environments”. Proc. of the 6th International Conference on Advances in Computer-Human Interactions (ACHI 2013), IARIA, 2013. / in press /

[9] D. M. Beder, A. C. da Silva, J. L. Otsuka, C. G. Silva, and H. V. da Rocha. “A Case Study of the Development of e-Learning Systems Following a Component-based Layered Architecture”. Proc. of the 7th IEEE International Conference on Advanced Learning Technologies (ICALT 2007), IEEE Press, Jul. 2007, pp. 21-25, doi: 10.1109/ICALT.2007.4.

[10] S. L. Oviatt. “Advances in Robust Multimodal Interface Design”, in IEEE Computer Graphics and Applications, vol. 23, no. 5, Sep. 2003, pp. 62-68, doi: 10.1109/MCG.2003.1231179.

[11] L. Nigay and J. Coutaz. “A Generic Platform for Addressing the Multimodal Challenge”. Proc. of 13th Conference on Human Factors in Computing Systems (SIGCHI 1995), ACM Press / Addison-Wesley Publishing Co., May 1995, pp. 98-105, doi: 10.1145/223904.223917.

[12] T. Mayes. “The ‘M’ Word: Multimedia interfaces and their role in interactive learning systems”, in Multimedia Interface Design in Education, A. D. N. Edwards and S. Holland, Eds. Berlin: Springer-Verlag, 1992, pp. 1-22, doi: 10.1007/978-3-642-58126-7_1.

[13] S. Bangalore and M. Johnston. “Robust Understanding in Multimodal Interaction”. Computational Linguistic, Cambridge, Massachusetts, vol. 35, n. 3, Sep. 2009, pp. 345-397.

[14] B. Dumas, D. Lalanne, and S. Oviatt. “Multimodal Interfaces: A Survey of Principles, Models and Frameworks”, in Human-Machine Interaction, D. Lalanne and J. Kohlas, Eds. Berlin: Springer Berlin / Heidelberg, 2009, pp. 3-26, doi: 10.1007/978-3-642-00437-7_1.

[15] J. Bouchet, L. Nigay, and T. Ganille. “ICARE: Software Components for Rapidly Developing Multimodal Interfaces”. Proc. of 6th International Conference on Multimodal Interfaces (ICMI 2004), ACM Press, Oct. 2004, pp. 251-258, doi: 10.1145/1027933.1027975.

[16] C. Duarte and L. Carriço. “A Conceptual Framework for Developing Adaptive Multimodal Applications”. Proc. of 11th International Conference on Intelligent User Interfaces (IUI 2006), ACM Press, Jan. 2006, pp. 132-139, doi: 10.1145/1111449.1111481.

[17] A. Stanciulescu, Q. Limbourg, J. Banderdonckt, B. Michotte, and F. Montero. “A Transformational Approach for Multimodal Web User Interfaces based on UsiXML”. Proc. of 7th International Conference on Multimodal Interfaces (ICMI 2005), ACM Press, Oct. 2005, pp. 259-266, doi: 10.1145/1088463.1088508.

[18] A. T. Neto, T. J. Bittar, R. P. M. Fortes, and K. Felizardo. “Developing and Evaluating Web Multimodal Interfaces – A Case Study with Usability Principles”. Proc. of 24th ACM Symposium on Applied Computing (SAC 2009), ACM Press, Mar. 2009, pp. 116-120, doi: 10.1145/1529282.1529306.

[19] J. A. Larson. “Should You Build a Multimodal Interface for Your Web Site?”. Available at < http://www.informit.com/ articles/article.aspx?p=29024>. [retrieved: Jan. 2013]

[20] A. Gruenstein, I. McGraw, and I. Badr. “The WAMI Toolkit for Developing, Deploying, and Evaluating Web-Accessible Multimodal Interfaces”. Proc. of 10th International Conference on Multimodal Interfaces (ICMI 2008), ACM Press, Oct. 2008, pp. 141-148, doi: 10.1145/1452392.1452420.

[21] A. Zaguia, M. D. Hina, C. Tadj, and A. Ramdane-Cherif. “Using Multimodal Fusion in Accessing Web Services”, in Journal of Emerging Trends in Computing and Information Sciences, vol. 1, n. 2, Oct. 2010, pp. 121-137.

[22] J. Nielsen. Usability Engineering. EUA: Morgan Kaufmann, 1993. 362 p.

A.M.B.E.R. Shark-Fin: An Unobtrusive Affective Mouse

Thomas Christy
School of Computer Science
Bangor University
Bangor, United Kingdom
thomas@thomaschristy.co.uk

Ludmila I. Kuncheva
School of Computer Science
Bangor University
Bangor, United Kingdom
l.i.kuncheva@bangor.ac.uk

***Abstract*—Analysing, measuring, recognising and exploiting emotion is an attractive agenda in designing computer games. The current devices for imputing physiological modalities are usually awkward to wear or handle. Here we propose a Shark-fin Mouse which streams three signals in real-time: pulse, electrodermal activity (EDA, also known as galvanic skin response or GSR) and skin temperature. All sensors are embedded into a fully functional computer mouse and positioned so as to ensure maximum robustness of the signals. No restriction of the mouse operation is imposed apart from the user having to place the tip of their middle finger into the Shark-fin hub. Boundary tests and experiments with a simple bespoke computer game demonstrate that the Shark-fin Mouse is capable of streaming clean and useful signals.**

***Keywords*-Interaction device; Affective gaming; Physiological sensors; Biometric feedback; Emotion**

I. Introduction

The video game industry, formally known as the interactive entertainment industry, has enjoyed perpetual growth since its foundation [8]. Video games are now experiencing the same level of financial-investment as seen in the film industry. The video game industry is among the largest entertainment industry worldwide, taking an estimated global income exceeding $50 billion in 2011 [15]. As computer systems have developed more processing power, video games have become more realistic and accompanied by the ever growing use of complex interactive devices, ranging from vibrating controllers [22] to fully tactile haptic feedback [23].

However, video games require more than representational graphics and tactility to be engagingly realistic. Video game environments also require greater humanlike emotive attributes or interactions between players and video game characters. Artificial intelligence (AI) attempts to bring the individuality of emotive human characteristics to human computer interactions (HCI). It has long been asserted that emotions are an important part of the human psyche and play a vital role in human communications [6]. The field Affective Computing (AC) has seen a dramatic rise in popularity over the past decade [24] [4] [25] permeating various disciplines such as computer science, electronic engineering and psychology. Growing on this trend, Affective Gaming (AG) is receiving significant consideration from academic and industrial fields [20] [3] [2] [12]. However, there is still no consensus on the best modalities and methods to use to collect emotional data. Many modes have been considered for use in gaming environments, see table I.

Table I
AFFECTIVE GAMING MODALITIES AND THE CURRENT CONTRIBUTORS.

	Reference	Modalities	Game
[2]	Ambinder	HR, eye movement, EDA, EEG, pupil dilation, EOG, posture, gesture, voice, face expression, respiration	Left4Dead2 Portal2
[3]	Bonarini	EDA, HR, pressure, temperature, gyroscope	Racing game
[5]	Chanel	HR, EOG, GSR, EEG, respiration, temperature	Tetris
[7]	Drachen	EDA, HR	Prey, Doom3, Bioshock
[9]	Gilleade	HR	Action based
[10]	Hoogen	Control tilt, pressure	Racing game
[11]	Jannett	Time, eye movement	HalfLife
[12]	Jones	Vocal cue/pitch/intonation, speech rate/volume	HalfLife Mod
[16]	McQuiggan	HR, EDA	Treasure Hunt (Source)
[21]	Nacke	EDA, EMG	HalfLife2 Mod
[26]	Rani	HR, EDA	Pong
[27]	Saari	User control knobs	Generic
[29]	Sykes	Game pad pressure	Space Invaders
[30]	Tijs	HR, EDA, respiration	Pacman
[31]	Tognetti	EDA	Racing game

Abbreviated terms: Heart Rate (HR), Electrodermal Activity (EDA), Electrooculography (EOG), Electroencephalography (EEG), Electromyography (EMG).

Emotions can be detected using a myriad of sensors. These sensors can be broadly divided into two distinct groups: behavioural and physiological, see figure 1. Behavioural recognition systems typically use cameras, microphones, Human Computer Interaction (HCI), and re-

 ISBN: 978-1-61208-250-9

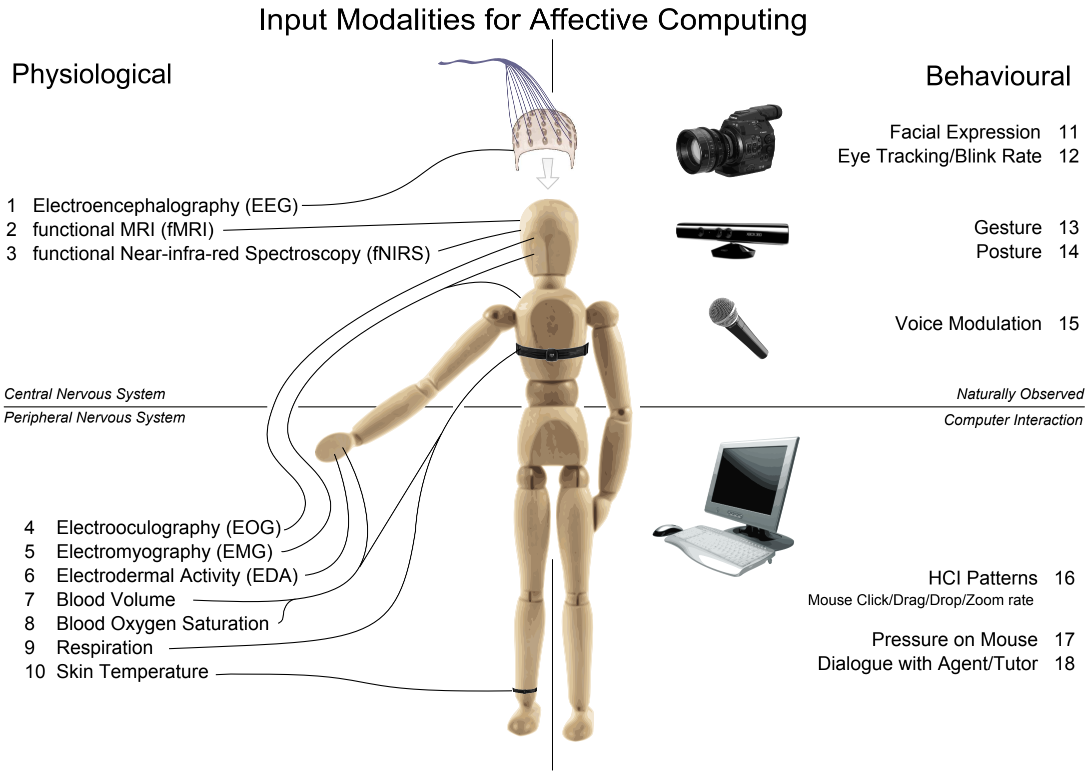


Figure 1. An illustration of the modalities for measuring emotion.

cently 3D depth sensors (Microsoft Kinect). Behavioural-based emotion recognition systems are generally unobtrusive; however, various facets of emotion are not always accurately displayed externally, particularly noted across different cultures [17]. Physiological patterns of emotion are less susceptible to cognitive artefacts; therefore give a truer representation of experienced emotion state [13]. Physiological recognition systems use various electrodes or sensors to detect changes in the body's peripheral nervous system [13]. However, physiological systems require sensors that are in direct contact with the user's skin or are attached to the user. This raises the questions of wearablity and comfort [3], particularly when introducing such a system or device to a gaming environment. Principally, video game input-devices are expected to add to the player's experience and not detract from the player's enjoyment [2].

For an affective interaction device to be adopted, it should be unobtrusive. The device should not hinder the player's enjoyment and should not handicap the player or the type of game that can be played. It should be intuitive to use. A player should not have to guess what the device should do. In addition, the set-up procedure, if any, should be short and straightforward. The player may get irritated by a delayed start of the game or by having to strap a device to their wrist, finger, chest of head.

One particular problem is the need to keep the physiological sensors still when in use, due to the high levels of amplification often needed to pick up signals from the body. Chanel et al [5] chose Tetris as a game that requires only one hand to play, so that the signals can be measured from the resting hand. Similarly, McQuiggan et al [16] relied on a static hand while the subject played Treasure Hunt, a bespoke HalfLife2-Source-engine driven game. However, having one hand tied in sensors and having to keep it still throughout the game may spoil the experience.

Many systems use sensors placed away from their respective amplifiers and analogue to digital converters; thus the signals are then carried through long (12"+) wires to be processed. Therefore, artefacts introduced due to cables-noise are amplified. To overcome this limitation, it would be advantageous to integrate the sensors and the amplifiers within a device already being utilised to control the game. Special care has to be taken when designing such a device, to ensure that the movements of the hand controlling the device are not affecting the sensor outputs.

During the 2011 Game Developers Conference, Mike Ambinder [2] highlighted the difficulty of obtaining a reliable pulse signal when actively playing a computer game. Obtaining a robust heart rate signal by using a photoplethysmograph (PPG) is challenging as it is highly sensitive to

 ISBN: 978-1-61208-250-9

movement artefacts. For this reason, to apply a PPG to a game controller, a mechanism is required to hold the sensor/finger still during use.

Research carried out by Bonarini et al [3] on a novel device attached to the player's head revealed several additional problems. Participants reported that the device stimulated excessive sweat, took too long to attach and was not very intuitive to use.

Taking all into account, we believe that there is a niche for a simple and unobtrusive device which can deliver some of the most important physiological modalities for emotion recognition in real time. This paper proposes such a device in the form of a computer mouse. Our device attempts to include pulse rate through a "shark fin" hub implemented as a small alteration of a standard computer mouse, an electrodermal activity (EDA) sensor and a infrared temperature sensor.

The remainder of this paper describes the basic AMBER Shark-fin Mouse set-up, together with the individual signals used. We review how each signal was tested for accuracy, both individually and as a complete system. The experimental set-up and video-game is discussed, followed by the results and finally our conclusion.

II. The Shark-fin Mouse

The humble mouse was first introduced as a prototype as early as 1963, by Douglas Engelbart and Bill English of the Stanford Research Institute. Since its inception it has formed the backbone of all HCI. The concept of an emotional mouse has been explored before [14]. However, limitations of the designs discussed in the Introduction hampered successful adoption of such devices.

The Advanced Multi-modal Biometric Emotion Recognition (A.M.B.E.R.) project was developed to create a low-cost, unobtrusive, physiology sensitive system, incorporating a minimal set of physiological sensors with a high potential for affect detection. The AMBER system also incorporates EEG and combines both hardware and software algorithms that support affect detection. The AMBER Shark-Fin biometric mouse, being part of the AMBER project, is a fully functional computer mouse which can output three physiological signals in real time. It allows biometric data to be collected as the player naturally participates in a video game. This allows the device to seamlessly blend into the PC gaming environment.

The AMBER prototype produces two digital data-streams. The first is the mouse position data and the second is the physiological affective data. Both mouse and data signals can be read by a single computer or split into two separate computer systems, for research purposes. For this paper, one computer was used. To enable a mouse-driven emotion interaction device to be functional, a novel hardware approach and design was needed. Preliminary testing demonstrated that the middle finger was the optimum choice for a strong blood flow volume needed by the photoplethysmograph (PPG) heart rate sensor. We observed that resting the wrist on a desk while holding the mouse decreases the blood flow in the middle finger. It was observed that the amplitude of the peaks in the pulse signal fell to a quarter of their initial values. To overcome this issue, the mouse was designed so as to provide support for the wrist (figure 2).

To maintain the mouse's full functionality, the scroll-wheel was moved from the centre finger (now being used for a pulse) to the right or left side (depending on preference), to be operated by the thumb, as shown in figure 2. To assist in the stabilisation of the middle finger, and to house both active component of the PPG, an arched housing was designed in the centre of the mouse, in place of a standard scroll-wheel. This also created a dark chamber to comply with the PPG's optical requirements.

The EDA sensors required the player to place a continuous pressure on the electrodes in order to prevent artefacts. Typical EDA devices place the electrodes between two fingers of the hand. Since the fingers move during clicking the mouse, the sensors were designed to connect with the palm where the least amount of movement was likely. Two 1cm x 1cm square electrodes were situated 1cm apart on the back of the mouse casing (figure 2).

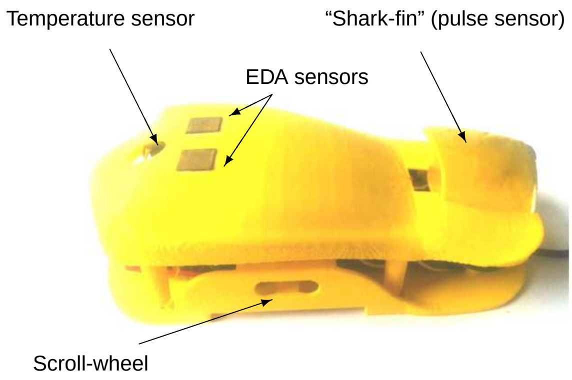


Figure 2. AMBER Shark-fin Mouse

To take account of any heat build-up invoked by thermal contact, a contactless IR thermometer was used. The temperature sensor was situated where it would receive a continuous and steady air convection from the hand. To assist further in heat build-up removal, the mouse casing was designed to be open, allowing air to circulate the circuitry, and the heat to dissipate naturally.

A. Signals

Many studies confirm the importance of combining the use of several physiological sensors to detect affect in video games, as seen in table I.

 ISBN: 978-1-61208-250-9

The variation in heart rate is a good indicator of stress and anxiety [13]. Using optical volumetric sensors known as a photoplethysmograph, the waveform of the heart's pulse can be detected through a thin part of the body, typically the finger or ear. The received signal produces a sharp peak in the waveform for every beat of the heart. Shelley and Shelley [28] demonstrate the wealth of information available through a photoplethysmograph sensor.

Electrodermal Activity (EDA), also referred to as Galvanic Skin Response (GSR), Skin Conductance Response (SCR) or Psycho Galvanic Reflex (PGR), measures the variance in electrical conductivity through the surface of the skin. EDA readings are effected through the sympathetic nervous system, making it a good indicator of stress and anxiety. EDA suffers with latency, with a delay of approximately one second for a response to be evoked, followed by approximately three seconds for the effect to dissipate. It is among the most basic and low cost physiological modalities available, and is widely used in physiological emotion recognition, including video games (table I).

Body temperature is affected by emotion, specifically joy, anger and sadness [18][19], and has been used for emotion recognition in video games. Temperature sensors fall into two general types: contact and non contact. Both types are sensitive to movement, which can introduce inaccuracies in the data collected. Movement is an important issue in the process of an active game play; hence, the positions of the sensors have to be chosen carefully.

Temperature is measured in Fahrenheit. Other units of measure are 3.3 volts used for both EDA and pulse data, capped by rapid-prototyping board FEZ-mini; measuring 1000 units per volt. Thus the minimum value will always be a positive integer or zero and the maximum value cannot exceed the integer value of 3,300. We avoid the typical contamination and artefacts through using capacitors in a smoothing circuit, as well as short cables and on-board computation.

III. Evaluation

Prior to running experiments, each component was individually evaluated to measure its performance against the expected output values. Each component was tested for minimum and maximum range and temporal responsiveness.

1) Temperature: The temperature sensor was expected to accurately respond to variances in the temperature of the skin. To test the range of the sensor, three objects with extreme variations in temperature where presented to the IR thermometer's $90°$ field-of-view (FOV), at an approximate distance of 5.5mm. The objects used were ice, a hot coffee pot and a human hand. The ice was taken from the freezer drawer of a fridge. Due to the reflective properties of ice, the ice-block was masked with a thin layer of paper. This had the expected effect of marginally raising the surface temperature. The same approach was used on the coffee-pot, again slightly altering the temperature. However, these alterations did not detract from the aim of the test, which was to measure the response of the thermometer to temperature changes. The objects where presented one after the other to the IR-thermometers FOV. The system collected the temperature data for 10 seconds and then paused for a key press; while the next object was selected. The pause was introduced to eliminate any danger when moving the hot-coffee-pot, while maintaining a 30 second data window. During this pause, the next object was presented, then the data recording resumed with key press.

After all three objects had their temperature read, the data was analysed, for range accuracy. Figure 3 shows the temperature range over the 30 second interval. The expected temperatures of all three objects can be seen in table II.

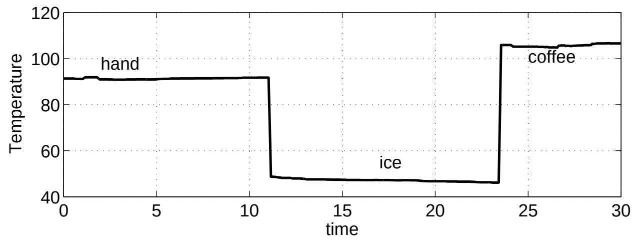


Figure 3. IR-Temperature 10s hand, 10s ice, 10s coffee

Table II
TEMPERATURE RANGE (°F)

Object	Minimum	Maximum	Mean observed
Hand	55	99	91.5
Ice (surface)	-30	50	48.7
Coffee	68	170	105.5

2) Photoplethysmograph: The photoplethysmograph sensor comprises of a 3mm infrared (IR) light emitting diode (LED) and a 3mm phototransistor. The sensor is required to optically detect blood volume, entering and leaving the middle finger. To test the functioning range of the sensor, a dense light blocking/absorbing card was used to completely block the light emanating from the IR-LED. When in situ, no light was able to enter the phototransistor and therefore no current should flow through the phototransistor. When the card was removed, the full range of light was passed to the phototransistor and the maximum current should be produced. The minimum and maximum voltage levels are 0-volts and 3.3-volts, respectively. The 3.3-volt is governed by the 3.3-volt analogue input cap of the devices circuitry. The card was placed in front of the sensor for approximately 5 seconds then removed for 5 seconds repeatedly, for a total duration of 30 seconds. Figure 4 depicts the recorded data.

 ISBN: 978-1-61208-250-9

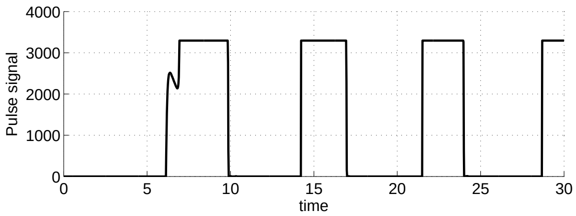


Figure 4. Heart Rate, 30 sec with 5sec covered 5sec uncovered repeatedly

3) Electrodermal Activity (EDA): To assess the function of the EDA sensors, a simple hand test was conducted. The palm of the right hand was placed on and off the device making contact with the two sensors, at intervals of approximately five seconds. The test began with the sensor untouched. This procedure was repeated for 30 seconds. Figure 5 demonstrates the unloaded (untouched) and loaded (touched) EDA circuit and the maximum and minimum range; 2552 and 196 respectively. This demonstrated the system's ability to respond during natural use.

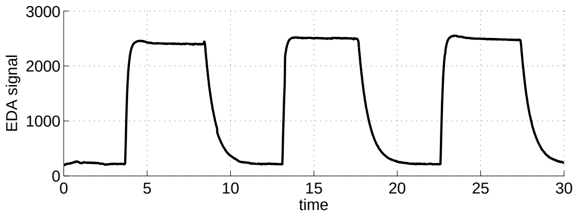


Figure 5. Electro-dermal activity, 30 sec with 5sec off 5sec on repeatedly

4) Complete system test: After ascertaining the minimum and maximum ranges, a trial was made to determine the ability of the sensors at detecting all signals together when applying the hand onto the device. For this test, all three sensors were recorded without any contact for approximately 10 seconds. After that, the right hand was placed on the mouse, covering both EDA and temperature sensors with the palm of the hand, and the middle finger was placed inside the photoplethysmograph cover. Figure 6 demonstrates the system's ability to pick up all three signals simultaneously. As expected the heart signal takes several seconds for the capacitors driving the amplification to find the pulse. After that, the pulse is cleanly detected. The EDA responded as expected with a small rise in current as soon as the sensors made contact with the skin. Similarly, the temperature sensor detected the change in temperature, from room temperature to skin temperature, almost immediately.

IV. Experimental Video-Game

After completing the individual sensor function tests, a simple video game was created to test the device's ability to detect basic affect in a gaming environment. The game involved moving the mouse over a randomly appearing white

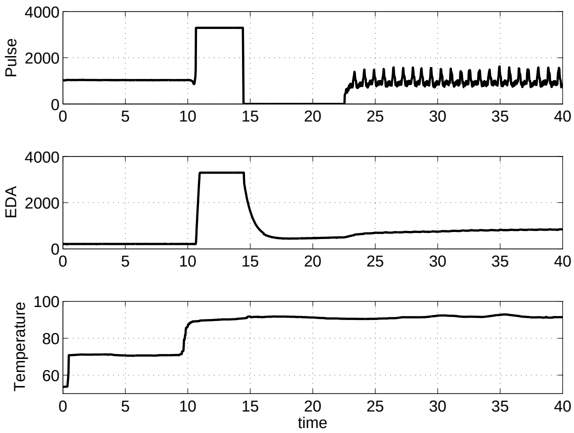


Figure 6. Heart rate (top), EDA (middle), Temperature (Bottom)

square, while avoiding a black square appearing in a straight path to the white square. The game set-up and the layout of the screen are shown in Figures 7 and 8, respectively. The game was specifically developed to invoke mild levels of stress. This was done by limiting the player's ability to freely move the mouse, while at the same time maintaining a need to move the mouse as quick as possible to gain a higher score within the countdown period. To achieve this, a mouse speedometer was created that limited the player's ability to score. If the mouse was moved at normal speeds, the scoring potential was reduced. If the players move the mouse quickly the energy level was reduced. If the energy bar became depleted, one of three lives was lost (Figure 8). The game ended after the counter reached zero or all three lives were lost.

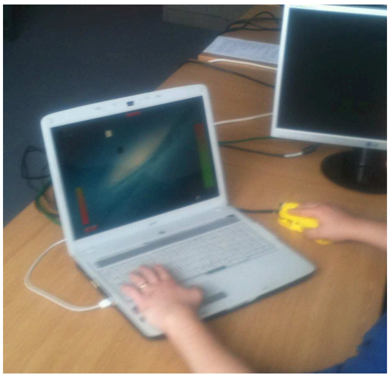


Figure 7. Experimental video-game layout

The data was recorded using two files. The first stored the computer game event data. The second recorded the biometric sensor data. For each file an epoch time-stamp was stored along with each iterative data point to enable temporal comparisons of the game-play and biometric data. To begin the experiment, the player placed their hand on the mouse and were given the rules of the game. The game commenced

 ISBN: 978-1-61208-250-9

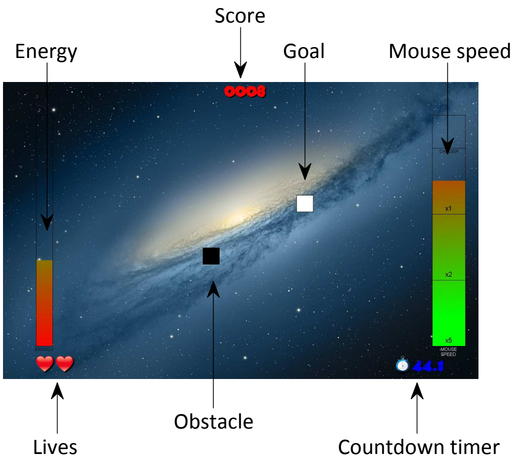


Figure 8. A screen shot of the experimental video-game.

and ran for a total of 90 seconds. During game-play, the state-of-play and game triggered-events were recoded to a text file, approximately every tenth of a second. Similarly the raw signals from the biometric sensors were collected through MATLAB along with the epoch time stamp, at approximately two tenths of a millisecond.

V. Results

The results of the trial highlight the effectiveness of the device at measuring clean physiological signals from a fully functional mouse, while playing a video-game. Figure 9 demonstrates the quality of the signals achieved during a live game trial. To test the potential of the proposed device

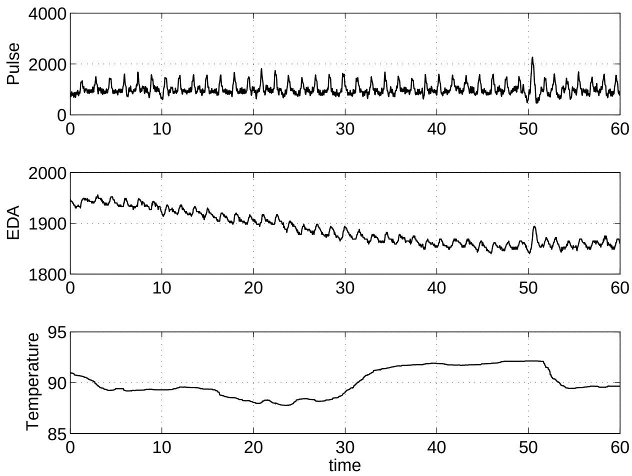


Figure 9. Typical physiological data output

at emotion recognition, we carried out off-line analyses of the data. The data consisted of the three variables (Pulse rate, GSR and skin temperature) measured along the whole approximately 90-second long game run. Each value in the sequence was the average of the previous 3 seconds of the signal. The pulse signal was transformed into pulse rate by the following steps:

1) The raw curve was smoothed with a window of 1/2 a second and then with a window of 1 second.
2) The locations of the peaks in the signal were identified.
3) The intervals between the subsequent peak locations was used to approximate the heart rate.
4) A linear interpolation was used to set the heart rate values between the peaks.

Two categories were formed. Assuming that the player is in state ‘Calm’ during most of the game run, we hypothesised that certain situations would provoke a negative state which we labelled ‘Agitated’. In this experiment, a state was labelled as ‘Agitated’ if all of the following held:

- The speed of the mouse exceeds a threshold of 200.
- The countdown clock indicates less than 25 seconds left.
- The player has lost at least one life thus far.

For example, game #3 produced 4129 data points, 56 of which were labelled ‘Agitated’ (3.78%), while the remaining points were labelled ‘Calm’. After removing the outliers (GSR signal less than 500 and temperature signal less than 50), and concatenating all 20 games, we obtained a labelled data set with 60,684 data points with 5332 data points (8.79%) in class ‘Agitated’. The task of developing a proper real-time classifier for such an imbalanced data set is one of our future lines of research. Here we are interested to find whether there are differences between the distributions of the two classes. A histogram was calculated for each of the three features and each class. The polygons of these histograms are shown in Figures 10, 11, and 12.

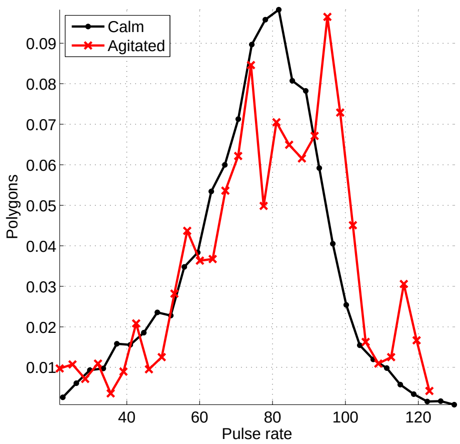


Figure 10. Polygons for classes ‘Calm’ and ‘Agitated’ for the Pulse Rate.

The polygon for class ‘Agitated’ has a more jagged appearance than the one for class ‘Calm’ because it was calculated from much fewer data points. More importantly, however, all distributions for class ‘Agitated’ are slightly

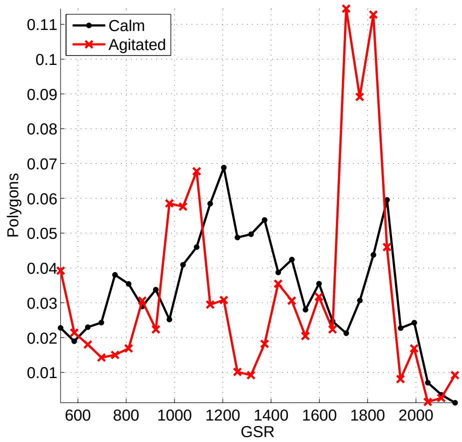


Figure 11. Polygons for classes ‘Calm’ and ‘Agitated’ for the GSR.

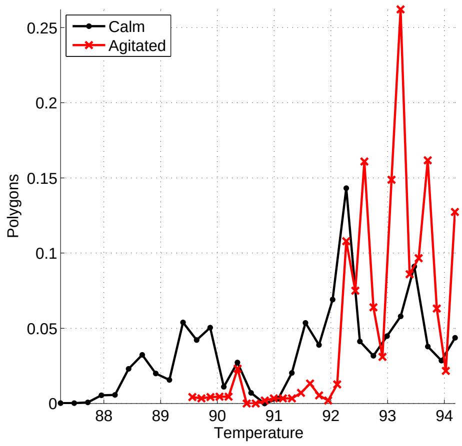


Figure 12. Polygons for classes ‘Calm’ and ‘Agitated’ for the Skin Temperature.

shifted to the right, indicating increased Pulse Rate, GSR and Skin Temperature. These findings are consistent with increased level of anxiety [13], which demonstrates the ability of the proposed device to output genuine and useful signals.

VI. Conclusion

In this study, we identified the need for a simple input device capable of streaming real-time physiological data. We proposed a novel solution, named the “Shark-fin Mouse” (due to the shape of the hub housing the plethysmograph). The mouse outputs three signals: pulse, electrodermal activity (EDA, often termed Galvanic Skin Response or GSR) and the skin temperature. Compared to the existing devices, the Shark-fin Mouse has the advantage of producing clean and reliable signals in real time in a subtle way. The Shark-fin mouse retains full functionality as a mouse, while hosting sensors positioned so as to ensure maximum robustness of the signals. Our tests and experiments demonstrated that the Shark-fin Mouse is capable of streaming the three physiological signals during a real game play.

The proposed device opens up interesting future research avenues. First, the Shark-fin Mouse was intended as a part of the A.M.B.E.R. project. The system will include additional affective modalities such as EEG and facial expression recognition. Second, we aimed at designing a device with a wide range of applications, including psychology, medicine, business, and entertainment industry. Data can be collected in different experimental environments. Due to the ability of the device to stream data in real time, experiments with emotional feedback can be carried out. The device can be turned into a useful gaming accessory because of its low cost and simplicity of concept and operation. It has to be mentioned that the Shark-fin Mouse is only a prototype at this stage. However, with the rise of the interest in affective computing we are planning to work towards wider academic and commercial deployment of the Shark-fin mouse.

Acknowledgment

This project has been supported by the Coleg Cymraeg Cenedlaethol www.colegcymraeg.ac.uk.

References

[1] M. Ambinder. Biofeedback in gameplay: How valve measures physiology to enhance gaming experience. Internet, 2011. Last access 24 Dec 2012.

[2] M. Ambinder. Biofeedback in gameplay: How valve measures physiology to enhance gaming experience. Game Developers Conference, 2011. Last access 24 Dec 2012.

[3] A. Bonarini, F. Costa, and M. Garbarino. Affective videogames: the problem of wearability and comfort. *SpringerLink: Human-Computer Interaction. Users and Applications*, 6764:649–658, 2011.

[4] R.A. Calvo and S. DMello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1):18–37, 2010.

[5] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*, pages 13–17. ACM, 2008.

[6] Charles Darwin. *The Expression of the Emotions in Man and Animals*. Murray, J., 1872.

[7] A. Drachen and A.L. Pedersen. Correlation between Heart Rate , Electrodermal activity and player experience in first-person shooter games. *Heart*, pages 49–54, 2009.

[8] Entertainment Software Association (ESA). Essantial facts about the computer and video game industry 2011. ESA E3 2011, 2011.

[9] K.M. Gilleade and J. Allanson. A Toolkit for exploring affective interface adaptation in videogames developing an affective videogame. *Proceedings of HCI International*, 2:370–374, 2003.

[10] V.D. Wouter M Hoogen, A.I Wijnand, and Y.A.W. de Kort. Effects of sensory immersion on behavioural indicators of player experience: movement synchrony and controller pressure. *digraorg*, 2009.

[11] C Jennett, A.L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9):641–661, 2008.

[12] C. Jones and J. Sutherland. Acoustic emotion recognition for affective computer gaming. *Affect and emotion in human-computer interaction: from theory to applications*, 4868:209, 2008.

[13] S. Kreibig. Autonomic nervois system activity in emotion: a review. *Biological psychology*, 84:394–421, 2010.

[14] Wenhui Liao, Weihong Zhang, Zhiwei Zhu, and Qiang Ji. A Decision Theoretic Model for Stress Recognition and User Assistance. *Systems Engineering*, pages 529–534, 2004.

[15] I. Livingstone. 'seismic shift' in computer education. BBC News, November 2011. Last access 24 Dec 2012.

[16] S.W. McQuiggan, S. Lee, and Lester J.C. Predicting user physiological response for interactive environments: An inductive approach. In *Proceedings of the 2nd Artificial Intelligence for Interactive Digital Entertainment Conference*, 2006.

[17] B. Mesquita, N.H. Frijda, and K.R. Scherer. Culture and emotion. *Handbook of cross-cultural psychology*, 2:255–297, 1997.

[18] H.G. Mittelmann, B. Wolff. Effective states and skin temperature: experimental study of subjects with "cold hands" and raynaud's syndrome. *Psychosomatic Medicine*, 1:271–292, 1939.

[19] H.G. Mittelmann, B. Wolff. Emotions and skin temperature; observations os patients during psychotherapeutic (psychoanalytic) interviews. *Psychosomatic Medicine*, 5(3):211–231, 1943.

[20] L E Nacke, S Stellmach, and C A Lindley. Electroencephalographic Assessment of Player Experience: A Pilot Study in Affective Ludology. *Simulation & Gaming*, 41(41):1–24, 2010.

[21] L.E Nacke and C.A. Lindley. Affective .ludology, flow and immersion in a first-person shooter: measurement of player experience. *Loading...: The Journal of the Canadian Game Studies Association.*, 3 (5)(5):21, 2009.

[22] Nintendo. Rumble pak. www.nintendo.co.uk, 2012. Last access 24 Dec 2012.

[23] Inc Novint Technologies. Novint falcon. www.novint.com, 2012. Last access 24 Dec 2012

[24] R.W. Picard. *Affective computing*. The MIT press, 2000. Last access 24 Dec 2012

[25] R.W. Picard. Affective computing : from laughter to IEEE. *Affective Computing, IEEE Transactions on*, 1(1):11–17, 2010.

[26] P. Rani. Maintaining optimal challenge in computer games through real-time physiological feedback mechanical engineering. *Pulse*, 2005.

[27] T. Saari, N. Ravaja, J. Laarni, and M. Turpeinen. Towards emotionally adapted games based on user controlled emotion knobs. *Human Factors*, 2005.

[28] K Shelley and S Shelley. Pulse oximeter waveform: photoelectric plethysmography. *Monitoring: Practical Applications for Anesthesia*, pages 420–428, 2001.

[29] J. Sykes and S. Brown. Affective gaming. *CHI 03 extended abstracts on Human factors in computing systems CHI 03*, page 732, 2003.

[30] T Tijs, D. Brokken, and W. Ijsselsteijn. Dynamic game balancing by recognizing affect. In *Proceedings of the 2nd International Confer- ence on Fun and Games.*, 2008.

[31] S. Tognetti, M. Garbarino, A. Bonarini, and M. Matteucci. Modeling enjoyment preference from physiological responses in a car racing game. *Computational Intelligence and Games (CIG)*, pages 321–328, 2010.