



# **ACHI 2012**

The Fifth International Conference on Advances in Computer-Human Interactions

ISBN: 978-1-61208-177-9

January 30- February 4, 2012

Valencia, Spain

## **ACHI 2012 Editors**

Leslie Miller, Iowa State University - Ames, USA

Silvana Roncagliolo, Pontificia Universidad Católica de Valparaíso, Chile

# ACHI 2012

## Forward

The fifth edition of The International Conference on Advances in Computer-Human Interactions (ACHI 2012) conference was held in Valencia, Spain, on January 30<sup>th</sup> – February 4<sup>th</sup>, 2012.

The conference on Advances in Computer-Human Interaction, ACHI 2012, 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 2012 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 2012 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 2012 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 2012. 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 2012 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 2012 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 the attendees enjoyed the beautiful surroundings of Valencia, Spain.

## **ACHI 2012 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 2012

## Committee

### ACHI 2012 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 2012 Technical Program Committee

Andrea Abate, University of Salerno, Italy  
V́ctor Acinas Garzón, Instalaciones Inabensa S.A. - Madrid, Spain  
Eugenio Aguirre Molina, University of Granada, Spain  
Pierre Borne, Ecole Centrale de Lille / IEEE Fellow, France  
Marius Brade, Technische Universität Dresden, Germany  
Luis Carriço, University of Lisbon, Portugal  
Ping Chen, University of Houston-Downtown, USA  
Ryszard S. Choras, University of Technology & Life Sciences - Bydgoszcz, Poland  
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, University of Bremen, Germany  
Carlos Duarte, University of Lisbon, Portugal  
Anastasios A. Economides University of Macedonia - Thessaloniki, Hellas  
Larbi Esmahi, Athabasca University, Canada  
Ben Falchuk, Telcordia Technologies, Inc., USA  
Elisabetta Farella, DEIS - University of Bologna, Italy  
Michael Farrugia, University College Dublin, Ireland  
Maryanne Fisher, Saint Mary's University - Halifax, Canada  
Karim Fraoua, Université Paris-Est- Conservatoire des arts et métiers, France  
Somchart Fugkeaw, Mahidol University - Nakhonpathom, Thailand  
Mathias Funk, Eindhoven University of Technology, The Netherlands  
Miguel García, Universidad Politecnica de Valencia, Spain  
Matjaz Garns, Josef Stefan Institute- Ljubljana, Slovenia  
Cristina Gena, Università di Torino, Italy  
Denis Gillet, Swiss Federal Institute of Technology (EPFL), Switzerland  
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

Lindsay D. Grace, Miami University, USA  
Vic Grout, Glyndwr University - Wrexham, UK  
Ido Guy, IBM Research-Haifa, Israel  
Till Røssvoll Halbach, Norwegian Computing Center / Norsk Regnesentral (NR), Norway  
Kristian J. Hammond, Northwestern University - Evanston, USA  
Jochen Hirth, University of Kaiserslautern, Germany  
Thomas Holz, University College Dublin, Ireland  
Clemens Holzmann, Upper Austria University of Applied Sciences, Austria  
Lichan Hong, Google Inc., USA  
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  
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  
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  
Satoshi Kurihara, Osaka University, Japan  
Andrew Kusiak, The University of Iowa, USA  
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  
Jean-Luc Lugin, Teesside University, UK  
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  
Deok Hee Nam, Wilberforce University, USA

Andreas C. Nearchou, University of Patras, Greece  
Radoslaw Niewiadomski , Telecom ParisTech, France  
Gregory O'Hare, University College Dublin (UCD), Ireland  
Fakri Othman, Cardiff Metropolitan University (UWIC), UK  
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  
Igor Petukhov, Blitzkrieg, Russia  
Marina Puyuelo Cazorla, Universitat Politècnica de València (UPV), Spain  
Didier Puzenat, Université des Antilles et de la Guyane - Pointe-à-Pitre, Guadeloupe  
Nitendra Rajput, IBM Research - New Delhi, India  
Xiangshi Ren, Kochi University of Technology, Japan  
Andreas Riener, University Linz - Linz/Donau, Austria  
Joel Rodrigues, Instituto de Telecomunicações / University of Beira Interior, Portugal  
Michael Rohs, Ludwig-Maximilians-Universität München, Germany  
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  
Mehmet Necip Sahinkaya, University of Bath, UK  
Munehiko Sasajima, Osaka University, Japan  
Dominique Scapin, INRIA - Le Chesnay, France  
Marc Seißler, University of Kaiserslautern, Germany  
Ross Smith, University of South Australia, Australia  
Jaka Sodnik, University of Ljubljana, Slovenia  
Weilian Su, Naval Postgraduate School - Monterey, USA  
Mauro Teófilo, Nokia Technology Institute, Brazil  
Daniel Thalmann, VRLab/EPFL - Lausanne, Switzerland  
Nikolaos Tsagarakis, Istituto Italiano di Tecnologia (IIT-Genova), Italy  
M. Sarosh Umar, Aligarh Muslim University, India  
José Valente de Oliveira, University of Algarve, Portugal  
Charlotte Wiberg Umeå University, Sweden  
Yingcai Xiao, University of Akron, USA  
Fan Yang, Nuance Communications, Inc., USA  
Jie Yin, CSIRO ICT Centre, Australia  
Claudia Zapata del Rio, Pontificia Universidad Católica del Peru - Lima , Peru  
Leyla Zhuhadar, Western Kentucky University, USA

## Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

## Table of Contents

GPU Based Burning Process Simulation <i>Ran Jiao, Liu Yonggan, and Hao Aimin</i>	1
Context-dependent Action Interpretation in Interactive Storytelling Games <i>Chung-Lun Lu and Von-Wun Soo</i>	7
Developing User-Centered Video Game Concepts for Language Learning <i>Yorick Poels, Jan-Henk Annema, Bieke Zaman, and Frederik Cornillie</i>	11
Online Casinos: The Addiction Under Control Role of Web 2.0 and re-documentarisation <i>Karim Fraoua and Christian Bourret</i>	17
Evaluation of User Interface Satisfaction of Mobile Maps for Touch Screen Interfaces <i>Ya-Li Lin</i>	22
Tablet PCs – An Assistive Technology for Students with Reading Difficulties? <i>Andrea A. Gasparini and Alma Leora Culen</i>	28
Spatial Ability and Map-Based Software Applications <i>Michelle Rusch, Sarah Nusser, Les Miller, Georgi Batinov, and Kofi Whitney</i>	35
A Presentation Support System by Expanding Embodiment with a Mobile Touchscreen Device <i>Michiya Yamamoto, Taku Murabayashi, and Tomio Watanabe</i>	41
Designing Mobile Apps for Visually Impaired and Blind Users <i>Javier Sanchez and Joaquin Selva Roca de Togores</i>	47
Building Bridges Between Elderly and TV Application Developers <i>Jose Coelho, Carlos Duarte, Pedro Feiteira, Daniel Costa, and David Costa</i>	53
Analysis of Volumetric Tactile Symbols Produced with 3D Printing <i>Jaume Gual, Marina Puyuelo, and Joaquim Lloveras</i>	60
Ethnographic Examination for Studying Information Sharing Practices in Rural South Africa <i>Amandeep Dhir, Imad Moukadem, Nobert Jere, Puneet Kaur, Sari Kujala, and Antti Yla-Jaaski</i>	68
Accessibility Study of Rich Web Interface Components <i>Juliana Braga, Rafael Damaceno, Rodrigo Leme, and Silvia Dotta</i>	75
Virtualization Technology for Multi-display Systems	80



<i>Igor Petukhov, Lyudmila Steshina, and Ilya Tanryverdiev</i>	
A Generic Approach to Interactive University Timetabling <i>Michael Zeising and Stefan Jablonski</i>	84
The Utility of Controlled Vocabularies within Bookmark Management Tasks <i>Siu-Tsen Shen and Stephen D. Prior</i>	90
Designing Multi-Modal Map-Based Interfaces for Disaster Management <i>Volker Paelke, Karsten Nebe, Christian Geiger, Florian Klompmaker, and Holger Fischer</i>	95
Multilingual Ontology Alignment Based on Visual Representations of Ontology Concepts <i>Srdan Mihic and Dragan Ivetic</i>	101
TsoKaDo: An Image Search Engine Performing Recursive Query Recommendation Based on Visual Information <i>Lazaros Tsochatzidis, Athanasios Kapoutsis, Nikos Dourvas, Savvas Chatzichristofis, Yiannis Boutalis, and Konstantinos Zagoris</i>	106
Head Nod and Shake Gesture Interface for a Self-portrait Camera <i>Shaowei Chu and Jiro Tanaka</i>	112
Towards 3D Data Environments using Multi-Touch Screens <i>Francisco R. Ortega, Armando Barreto, Naphtali Rishe, and Malek Adjouadi</i>	118
User Interface for Trust Decision Making in Inter-Enterprise Collaborations <i>Puneet Kaur, Sini Ruohomaa, and Lea Kutvonen</i>	122
Human Operator Perspective to Autonomic Network Management <i>Marja Liinasuo, Iina Aaltonen, Hannu Karvonen, Beatriz Fuentes, and Alfonso Castro</i>	128
Contents Enforme: Automatic Deformation of Content for Multi-features without Information Loss <i>Hiroaki Tobita</i>	135
Authenticated Tangible Interaction using RFID and Depth-Sensing Cameras <i>Florian Klompmaker, Holger Fischer, and Helge Jung</i>	141
Generic Brain-computer Interface for Social and Human-computer Interaction <i>Julita de la Vega Arias, Christoph Hintermuller, and Christoph Guger</i>	145
A Conversational System to Assist the User when Accessing Web Sources in the Medical Domain <i>Marta Gatius and Tsetsegkhand Namsrai</i>	150
Face Detection CUDA Accelerating	155

<i>Jaromir Krpec and Martin Nemeč</i>	
User Attention in Mobile Devices <i>Pekka Isomursu, Minna Isomursu, and Mari Ervasti</i>	161
Semiautomatic Evaluation of Websites Usability <i>Eva Garcia, Antonio Garcia-Cabot, Luis de-Marcos, Salvador Oton, and Jose-Ramon Hilera</i>	165
Motion-sound Interaction Using Sonification based on Motiongrams <i>Alexander Refsum Jensenius</i>	170
Touchscreen Interfaces for Visual Languages <i>Michael Hackett and Philip T. Cox</i>	176
Design Guidelines for Hybrid 2D/3D User Interfaces on Tablet Devices - A User Experience Evaluation <i>Katri Salo, Leena Arhipainen, and Seamus Hickey</i>	180
A Usage-Centered Evaluation Methodology for Unmanned Ground Vehicles <i>Jurriaan van Diggelen, Rosemarijn Looije, Tina Mioch, Mark Neerincx, and Nanja Smets</i>	186
An Error Analysis Model for Adaptive Deformation Simulation <i>Umut Kocak, Karljohan Lundin Palmerius, and Matthew Cooper</i>	192
On the Evaluation of Auditory and Head-up Displays While Driving <i>Christina Dicke, Grega Jakus, Saso Tomazic, and Jaka Sodnik</i>	200
Evaluating the Usability and the Communicability of Grid Computing Applications <i>Cristian Rusu, Silvana Roncagliolo, Arturo Figueroa, Virginica Rusu, and Dorian Gorgan</i>	204
Practical Usability in XP Software Development Processes <i>Zahid Hussain, Martin Lechner, Harald Milchrahm, Sara Shahzad, Wolfgang Slany, Martin Umgeher, Thomas Vlk, Christina Koeffel, Manfred Tscheligi, and Peter Wolkerstorfer</i>	208
Emotion Recognition Through ANS Responses Evoked by Negative Emotions <i>Eun-Hye Jang, Byoung-Jun Park, Yeongji Eum, Sang-Hyeob Kim, Chul Huh, and Jin-Hun Sohn</i>	218
Identification of Optimal Emotion Classifier with Feature Selections Using Physiological Signals <i>Byoung-Jun Park, Eun-Hye Jang, Sang-Hyeob Kim, Chul Huh, and Jin-Hun Sohn</i>	224
Interacting with Navigation Devices: A Case Study <i>Javier Calle, Esperanza Albacete, Enrique Sanchez, and Dolores Cuadra</i>	230
Kalman Filter for Tracking Robotic Arms Using low cost 3D Vision Systems	236

*Enrique Martinez-Berti, Antonio-Jose Sanchez-Salmeron, and Francesc Benimeli*

Predicting Performance and Situation Awareness of Robot Operators in Complex Situations by Unit Task Tests 241  
*Tina Mioch, Nanja J. J. M. Smets, and Mark A. Neerincx*

Evaluation of Cognitive Effort in the Perception of Engineering Drawings as 3D Models 247  
*Florin Girbacia*

A Recommendation Method Based on Contents and User Feedback 251  
*So Ryoung Kim, Sang Min Choi, Yo Sub Han, and Lae Hyun Kim*

Developing a Human Computer Interaction Course for an Information Technology Major 256  
*Cynthia Lester*

Physical Instructional Support System Using Virtual Avatars 262  
*Tomoaki Ogawa and Yasushi Kambayashi*

Situated Cognitive Engineering: The Requirements and Design of Automatically Directed Scenario-based Training 266  
*Marieke Peeters, Karel van den Bosch, John-Jules Ch. Meyer, and Mark A. Neerincx*

Augmented Reality in Minimally Invasive Surgery 273  
*Lucio Tommaso De Paolis De Paolis and Giovanni Aloisio Aloisio*

Increased Cognitive Load in Resolution of Problems Caused by Human Error on New Aircrafts 278  
*Edgard Martins*

Using Social Media for Collaborative Learning in Higher Education: A Case Study 285  
*Na Li, Sandy El Helou, and Denis Gillet*

AlgoPath: A New Way of Learning Algorithmic 291  
*Estelle Perrin, Sebastien Linck, and Frederic Danesi*

Alleyoop: Interactive Information Retrieval System with Sketch Manipulations 297  
*Hiroaki Tobita*

Interactive Hand Gesture-based Assembly for Augmented Reality Applications 303  
*Rafael Radkowski and Christian Stritzke*

A Dance Training System that Maps Self-Images onto an Instruction Video 309  
*Minoru Fujimoto, Tsutomu Terada, and Masahiko Tsukamoto*

Interaction in Augmented Reality by Means of Z-buffer Based Collision Detection 315

*Yasuyuki Souma, Hidemi Yamachi, Yasuhiro Tsujimura, and Yasushi Kambayashi*

The Effect of Metacognition in Cooperation on Team Behaviors 319  
*Kohei Nonose, Taro Kanno, and Kazuo Furuta*

Educational Playability Analyzing Player Experiences in Educational Video Games 326  
*Amer Ibrahim, Francisco Luis Gutierrez Vela, Jose Luis Gonzalez Sanchez, and Natalia Padilla Zea*

User Experience: Buzzword or New Paradigm? 336  
*Dominique Scapin, Bernard Senach, Brigitte Trousse, and Marc Pallot*

Enhancing Automatic Detection of Frustration Induced During HCI with Moment-based Biosignal Features 342  
*Dimitris Giakoumis, Dimitrios Tzovaras, and George Hassapis*

Adaptivity Considerations for Enhancing User-Centric Web Experience 348  
*Marios Belk, Panagiotis Germanakos, Panagiotis Zaharias, and George Samaras*

Modality Preferences of Different User Groups 354  
*Benjamin Weiss, Sebastian Moller, and Matthias Schulz*

Users' Trust and Secure Feeling towards Cloud Services 360  
*Kaarina Karppinen*

# GPU Based Burning Process Simulation

Ran Jiao, Liu Yonggan, Hao Aimin

State Key Laboratory of Virtual Reality Technology and Systems, Beihang University  
Beijing, China  
E-mail: ranjiao@gmail.com

**Abstract**—We present a method of simulating the process of burning phenomena on generic polyhedral objects. By mapping the object's surface to a 2D space, the fire front expansion can be calculated efficiently on GPU (Graphics Processing Unit). The state of decomposition is updated according to the fire front and the consumption of solid fuel. During the simulation loop, both the fire front and the solid fuel consumption are updated respectively. In order to achieve a better performance, most routines of the simulation are processed on GPU. The entire simulation could run at an interactive rate on a normal PC.

**Keywords**-burning; fire spreading model; deformation.

## I. INTRODUCTION

Combustion is an important natural phenomenon which is widely used in virtual environment, such as video games, industrial simulations, etc. Although lots of researches are focused on fluid simulation, such as water, fire and smoke, most of these works are dedicated to fluid simulation itself. Little attention was paid to fire propagation and the objects being burnt. In today's game development, there's still no usable technique that can simulate solid combustion. However, real-time simulation of object's combustion process is increasingly attracting more attention, since it can dramatically improve the quality of fire simulation. Although physics based approach can produce convincing result [1], it's still too expensive for real-time application. Most of modern game engines like CryEngine 3 (game engine released by Crytek) and Unreal Engine 3 (a widely used computer game engine developed by Epic Games) are still using simple deformation animation [2] and particle system approach [3], which end up with coarse results and requiring artists to tune every burning solid.

While trying to simulate the process of burning, many physical concepts are involved. Firstly, the material and the geometrical structure of the object being burnt can affect the speed of fire propagation, which will affect the burning process significantly. Secondly, along with the combustion, the object will lose combustible stuff and decompose, which will change the shape of the object and have an influence on the fire propagation as well.

We demonstrate a real-time simulation method of burning phenomena on generic polyhedral objects. The main contributions of this paper are as follows:

- A GPU based algorithm to model the expansion of fire on polygon surface.

- Introduce a modified mass-string model to describe the physical deformation of a decomposing and burning mesh.
- Raise a method which can detect polygon self-intersection so as to avoid incorrect rendering results after topological structure changes.

Most routines of this method could be done on GPU. The burning state is stored as textures and additional vertex information in video card's memory. Since all data structure is stored in textures, nearly all the calculation is able to be done on GPU. The simulation quality of fire expansion could be adjusted by simply modifying the resolution of the burning state texture. The topological structure changes of the object during the combustion are updated correctly. Meanwhile, the accuracy of the simulation is not affected at all. This method can be applied easily to most polygon meshes.

The next section presents some related work about deformation, fire simulation and combustion simulation. Section III introduces a model simulation algorithm, describing how fire spreads on a mesh. There we will describe our method and how to process the whole algorithm on GPU. Section IV describes the physics model calculating the deformation of a burning mesh. The overall performance is excellent since the algorithms of both fire propagation and mesh deformation are calculated efficiently on GPU. Final results are presented in Section V.

## II. RELATED WORK

### A. Deformation

Free-form deformation (FFD) is widely used in both commercial software such as 3D Studio Max and real-time simulation as an important tool for computer-assisted geometric design and animation. FFD is firstly developed by Barr [1] and later improved by Sederberg and Parry [4]. A generic approach is proposed by Milliron et al [5].

Recently, some Laplacian coordinates-based deformation methods like VGL [6] are developed to achieve better results for large scale deformation.

### B. Fire Simulation

The early models of fire simulation are mostly based on particle system firstly developed by Reeves [7]. Although article system is widely used in real-time rendering such as video games, the result it produces is still not convincing.

In order to achieve realistic and physics correct rendering result, physics based fire simulation is developed by using the algorithm of Computational Fluid Dynamics, CFD [8]. Jos Stam introduced SPH into computer graphics in [9], and later he proposed an unconditionally stable model to solve the Navier-Stokes equation in Stable Fluids [1], making physics based fluid simulation less expensive. Although physics based fluids simulation has been studied for over 20 years, its calculation is still too expensive for real-time rendering on current PC.

### C. Combustion Simulation

Konrad Polthier and Markus Schmies [10] modified geodesic flow method to make it able to work on polyhedral surface. Reference [10] computes the evolution of distance circle on polyhedral surface and develops a method to visualize the geodesic circle. This method is used in [11] to simulate the fire front.

Hauyoung Lee and Laehyun Kim [11] used geodesic flow method to simulate the combustion process on polyhedral surfaces. In [11], fire fronts are evolved directly on the surface of arbitrarily complex objects by using modified geodesic flow method. Wind field model is also used to achieve animator control and motion complexity. Combustion process is treated as the propagation of fire front only, which is suitable to simulate fire spread on terrain, but cannot achieve complex results. Besides, this method did not take the decomposition of burning objects into consideration, which is necessary for most combustible materials.

Zeki Melek and John Keyser presented a burning objects simulation framework [12]. They simulated flame and solid separately, and used a heat transfer mechanism to transport energy between two systems. In flame simulation phase, a modified version of Stable Fluids [8] approach is used. The fluid solution is applied to fuel gas, exhaust gas and heat. In solid simulation phase the burning boundary is firstly computed by using level set method [13]. Then, a regular 3D grid is used to represent the decomposition of the solid and how much fuel left in the solid. If any part of the solid reaches the ignition temperature, it will start burning and release fuel to flame system.

Singguang Liu et al. proposed a unified framework [14] for simulation burning phenomena of thin-shell objects such as paper, cloth, etc. Fire spreading is modeled as burning state propagation on NxN cells of a 2-dimension space. They also proposed a method used to calculate the deformation of the thin-shell object based on the state of combustion, and then apply deformation by using FFD. Since the simulation is processed by CPU, this method can only achieve 10 frames per second.

## III. FIRE SPREADING MODEL

### A. Mapping

The burning state is composed of two stages. Firstly, every vertex has its own vertex state which records the burning state, deformation state, etc. Secondly, for points inside a triangle, the point state is updated and stored in a texture, recording how much this point is burnt.

The burning state space is treated as a float texture in our implementation. In order to use a texture to describe the fire spreading, a mapping from  $R^3$  to  $R^2$  is involved.

Every single triangle of the burning mesh should have its unique state, which requires the texture mapping should not be overlapped. In order to meet this requirement, a little more care should be taken while creating texture mapping.

Since the mapping from  $R^3$  to  $R^2$  will create discontinuous area, artists should also be careful about the part of discontinuous mapping, because burning state updating requires the neighboring vertices are also connected in burning state map.

### B. Burning State

For a mapped point on burning state surface, it should be at one of the three states: normal, burning, and burnt. We define the burning state of point  $(x, y)$  at time step  $n$  as

$S_{xy}^n$ . The value of  $S_{xy}^n$  shows how much it has burnt. State value 0 means state normal, and state 1 means burnt. At the start of the simulation, at least one point on this surface is in burning state, and these burning points will ignite nearby points and spread the fire. If one point has been burnt for a certain time, it goes to state burnt. Burnt points will change its appearance and no longer get updated in fire front spreading and mesh deformation.

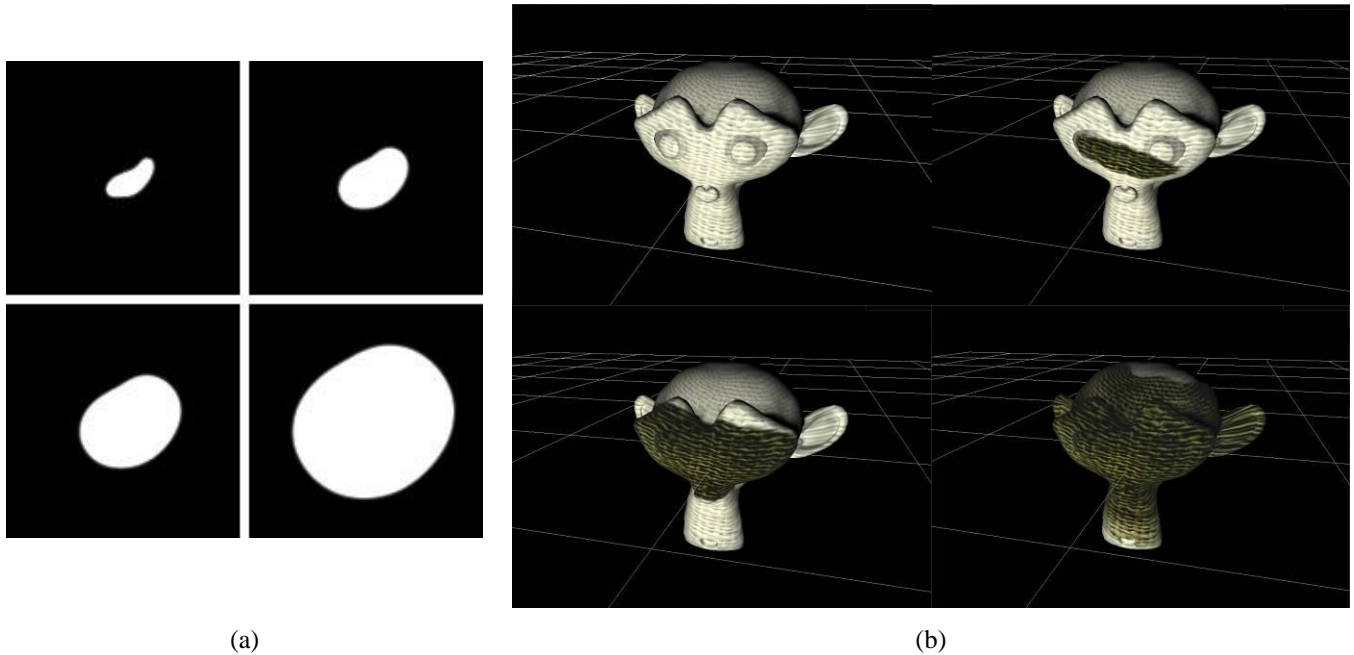
After mapping the mesh surface to a 2D surface, the fire propagation could be implemented as state changing on a texture.

For a burning state  $S_{xy}^{n+1}$  of point  $(x, y)$  at time step  $n+1$ , its state could be approximately evaluated by using the state  $S_{uv}^n$  of nearby points at time step  $n$ . The combustion state is updated each step by using a convolution in a region  $\sigma$  around the point  $(x, y)$ :

$$\begin{aligned} S_{xy}^{n+1} &= \int_{(u,v) \in \sigma} W(u,v) D(u,v) S_{uv}^n \\ &\approx \sum_{(u,v)} W(u,v) D(u,v) S_{uv}^n \end{aligned} \quad (1)$$

$W(u,v)$  is the weight function, which should be in inverse proportion to the distance between  $(x, y)$  and  $(u, v)$ .  $W(u,v)$  and area being summed up will affect the propagation speed. Each time we retrieve a value of  $S_{uv}^n$  a texture sampling will be invoke, since the burning state is stored in a texture. A much too big sigma area will significantly slow down this process. In our implementation we will sample 16 pixels around point  $(x, y)$ .  $D(x, y)$  is the propagation speed of fire front:

$$D(u, v) = C_{uv} C_D \cdot (1 + C_G \cdot \overrightarrow{N_{uv}} \cdot \overrightarrow{G}) \cdot \Delta t \quad (2)$$



(a)

(b)

Figure 1. The burning process on the burning state map. Burning state texture is showed in (a), the color of the texture represents how much it's burnt: black is not ignited yet, white is burnt. The burning speed and quality could be adjusted by using parameters. After mapping the burning state texture to the mesh, it could be used to calculate the color and vertex position, demonstrated in (b).

$\vec{G}$  is the direction vector of gravity, and  $\vec{N}_{uv}$  is the normal direction of point  $(u, v)$ .  $C_D$  is a constant parameter which controls the overall burning speed, and  $C_G$  is a constant parameter about how much the burning speed is affected by burning direction. We introduce gravity direction  $\vec{G}$  here for the reason that the flame is always burning against the direction of gravity, and the burning speed is also affected by the angles between surface normal and gravity direction.  $C_{uv}$  stands for the material multiplier for point  $(u, v)$ , which is stored in a texture to describe the burning speed for every point on the mesh surface.

As all calculation above is manipulating texture, it is easy to have this whole process implemented with GPU acceleration. This makes our implementation far more efficient than Liu's in [14].

#### IV. DEFORMATION

##### A. Modeling Deformation

The burning mesh is deformed by two forces: the string force  $F_S$  caused by nearby solid surface's deformation and the decomposition force  $F_D$  caused by the consumption of inner solid fuel:

$$F = \alpha F_S + \beta F_D \quad (3)$$

$\alpha$  and  $\beta$  are parameters of object material, which demonstrate how much these two forces should affect the

object. In our final result, we employ  $\alpha = 0.5$  and  $\beta = 3.0$ , making a wood like burning deformation.

$F_S$  is composed of structure force, shear force and flexible force [15]:

$$F_S = F_{\text{structure}} + F_{\text{shear}} + F_{\text{flexion}} \quad (4)$$

The consumption of solid fuel is calculated and stored while updating the burning state. Using this information of fuel consumption, we can calculate the force applied to the mesh caused by the decomposition.

Consider a small part of a burning mesh demonstrated in Figure 2. we can resolve the movement of vertex A caused by decomposition by calculating how much its volume changes:

$$\begin{aligned} \Delta B_A &= B_A - B_{A'} \\ &= D_C \cdot (V_{ABCDE} - V_{A'BCDE}) \\ &\approx D_C \cdot P_A P_{A'} \cdot S_{BCDE} \end{aligned} \quad (5)$$

The  $B_A$  represents how much the fuel is left,  $P_A$  and  $P_{A'}$  are the vertex's original and transformed position respectively.  $V_{ABCDE}$  and  $V_{A'BCDE}$  are the volume of the origin and transformed area.  $D_C$  is the density of the fuel, and  $S_{BCDE}$  is the size of polygon  $BCDE$ . The consumption of the solid fuel is approximately evaluated by measuring the distance the vertex has moved.

According to (5),  $P_A P_{A'}$  could be resolved:

$$|P_A P_{A'}| \approx \frac{\Delta B_A}{D_C \cdot S_{BCDE}} \quad (6)$$

By using the normal vector  $N_A$ , we can approximately get the offset vector:

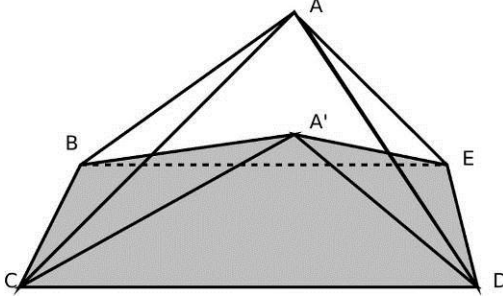


Figure 2. The deformation is calculated by evaluating how much solid fuel lost after vertex  $A$  moved to  $A'$  caused by decomposition.

According to the classic physical theory, we can get this equation:

$$\begin{aligned} F_D &= m \cdot a \\ dP_A &= \frac{1}{2} a \cdot dt \end{aligned} \quad (7)$$

Where  $m$  is the mass of mesh in Figure 2,  $a$  is the acceleration, and  $dt$  is the time step in our simulation. Then we can get  $F_D$ :

$$\begin{aligned} F_D &= m \cdot a \\ &= m \cdot \frac{2dP_A}{dt} \end{aligned} \quad (8)$$

In our implementation, the burning state is stored in a texture so that we can use GPU to calculate the fire front's expansion. So we need to sample the burning texture according to adjacent triangles' size.

Before each frame's deformation, we downsample the burning state map into  $n$  stages by halving the texture resolution. For each of these textures, the size a pixel represents is  $\frac{S}{w \cdot h} \cdot 2^n$ , and we can find a  $n'$  satisfying this equation:

$$\frac{S_{n'}}{w \cdot h} \cdot 2^{n'} \leq S_A k < \frac{S_{n'+1}}{w \cdot h} \cdot 2^{n'+1} \quad (9)$$

Where  $k$  is the how many pixel need to be sampled.  $k$  is the number of pixels we sampled in equation (1), in our implementation  $k = 16$ .  $w$  and  $h$  stand for the width and height of the burning state texture. By sampling the  $n'$ th burning state texture we can know how much fuel is used and get  $\Delta B_A$  for (6).

After getting both  $F_D$  and  $F_S$ , we can compute how much this vertex should move along its normal direction.

Since the area size is used in (7), we need to precompute the area size of the mesh before rendering, and pass it to the video card as vertex attributes. The normal vector of the vertex should also be updated as the faces are all deformed.

### B. Topological Changes

Along with the combustion process, the topological structure might change. We use a simple method based on video cards' graphical pipeline function to avoid incorrect rendering results.

Assume an object with two points  $P_1$  and  $P_2$ , which face the opposite direction. At the beginning, they do not cross each other, but after being burnt, the vertices will be pushed backwards according to their normal vector and burning state. On certain condition,  $P_1$  and  $P_2$  may get crossed as demonstrated in Figure 4. This will certainly ends up with incorrect rendering results. However, this could be fixed by adding a special pass to the rendering loop.

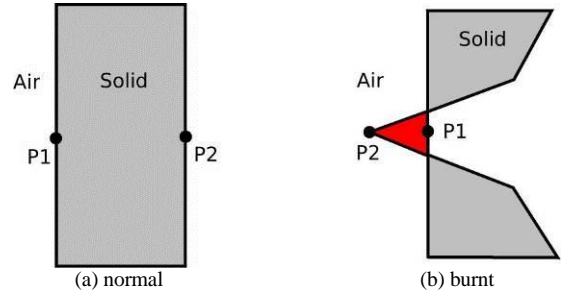


Figure 3. Combustion caused topological structure changes. Considering eyes looking from right to left,  $P_2$  in (a) should be the front face, and  $P_1$  on back face. After burnt  $P_2$  might goes to the back of  $P_1$  in (b). Red area in (b) will cause incorrect rendering with normal rendering routine.

A special Z-Pass and a depth test are introduced to eliminate incorrect rendering results caused by topological changes. In render loop, Z-Pass is used just after updating the burning state. Both front and back faces are rendered and have their depth written to a render target with depth test turned on. If a front face gets blocked by a back face just like what happens in Figure 4(b), the render target will have the back face's depth (it's also the nearer face) saved.

During the rendering pass the back face cull is turned on, and each vertex's depth value is compared to value stored in render target. Pixel in red area of Figure 4(b) will have further depth value than that we stored in Z-Pass, which means it has already crossed the face of another size, and this pixel will be discarded.

In our approach to cull incorrect pixels, self-collision detection is changed into a screen space problem. Since our technique used to detect topological changes can be implemented as a simple post-process, it won't be affected by mesh's vertices count and will run very fast on GPU.



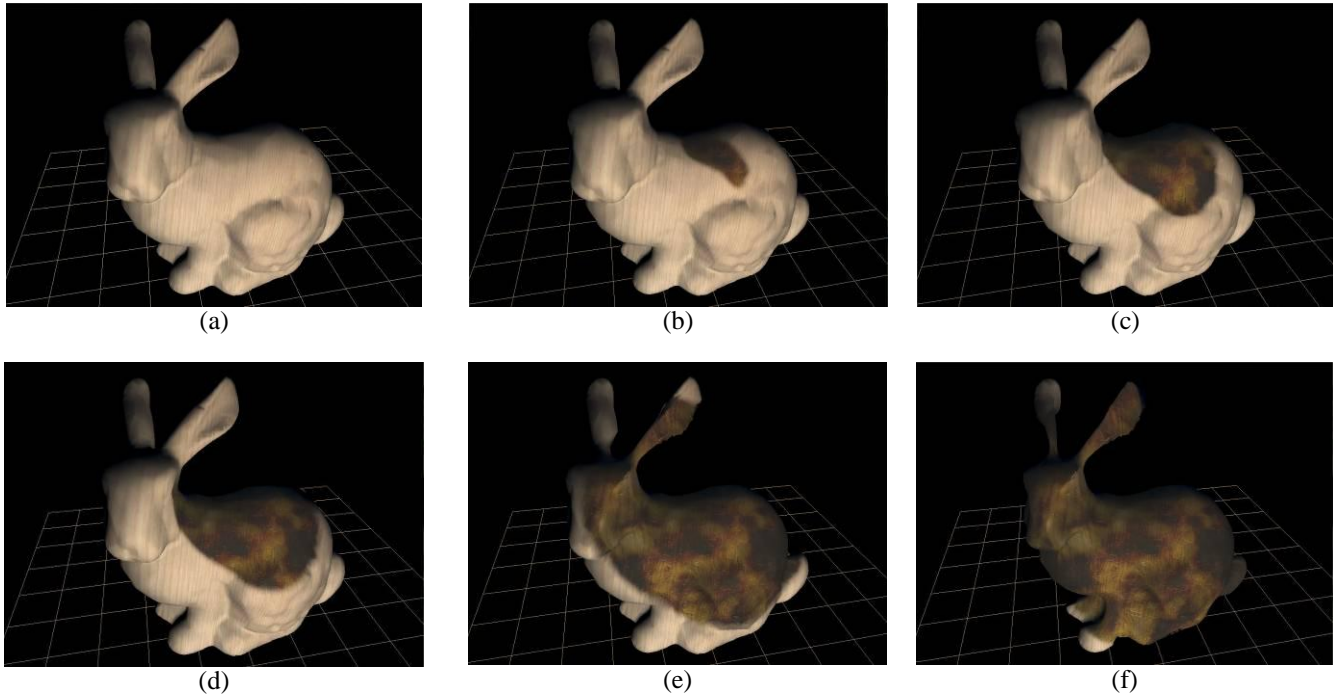


Figure 4. Combustion simulation of a bunny. The bunny mesh is composing of 4000 vertices and a 1024x1024 burning state texture.

V. RESULTS

Using the method we have introduced, we are able to simulate various kinds of burning meshes. The simulations are ran on Nvidia Geforce GT 425m, Intel Core i5 (2.53G Hz), 4G RAM and 1024x768 screen resolution. A 256 × 256 burning state texture and a mesh of 4000 vertices are used in scene and we got more than 100 fps. The result is simulated with  $\alpha = 0.2$ ,  $\beta = 1.0$ ,  $C_D = 0.3$ ,  $C_G = 0.1$ .

The whole process is shown in Figure 3. We initialize the burning state with a small area already burning on bunny’s back, and the fire spreads all over the mesh. Together with the burning process the mesh also shrink and deformed. You can see the fire combustion process is like the one in [14], but utilized on a normal solid mesh with much better performance.

TABLE I. PERFORMANCE

Burning State Texture Resolution	Mesh Vertices Count	Time Consumption(ms)		
		Fire Front	Self-Collision	Deformation
128x128	4000	2.4	4.9	1.4
	31000	2.4	4.9	1.6
256x256	4000	2.5	4.8	1.4
	31000	2.5	4.9	1.6
512x512	4000	2.5	4.8	1.4
	31000	2.5	5.0	1.6
1024x1024	4000	2.7	4.8	1.4
	31000	2.7	5.0	1.6

The performance with different state texture resolution and different mesh for algorithm described in this paper is listed in TABLE I. The performance of fire front and self-

collision calculation is almost not affected by mesh vertices count, because both of them are screen space algorithm. The overall performance benefits a lot from GPU’s parallel computing capability, it’s easy to render more than 100 frames per second.

VI. CONCLUSION AND FUTURE WORK

We have proposed a method of simulating combustion process of general polygon mesh, and it is able to be applied to most meshes. This method can run much more efficiently than similar methods in [11] and [13], since all calculation is done by GPU and only the deformation algorithm is affected by the scene’s complexity. In order to have topological structure change processed more elaborately, the mesh self-intersection could be calculated by using physics engines such as Nvidia PhysX, and recalculate the topological structure of vertices and triangles. Other deformation model could also be used to simulate the decomposing process of burning mesh. In our work we focused on fire expansion and mesh deformation rather than the fire simulation along with burning. In order to render fire during combustion process, the burning state could also be used for fire simulation. Wind field [16] could also be used to control the burning process and would produce more realistic result.

REFERENCE

- [1] J. Stam, "Stable fluids," in Proceedings of the 26th annual conference on Computer graphics and interactive techniques, New York, NY, USA, pp. 121-128, 1999.
- [2] S. F. Gibson and B. Mirtich, B. "A Survey of Deformable Models in Computer Graphics," Tech. Rep. TR-97-19, Mitsubishi Electric Research Laboratories, Cambridge, MA, November 1997.

- [3] N. Froster and R. Fedkiw, "Practical animation of liquids," Proceedings of SIGGRAPH 2001. New York, NY: ACM, pp 23-30, 2001.
- [4] T. W. Sederberg and S. R. Parry, "Free-form deformation of solid geometric models," in SIGGRAPH '86: Proceedings of the 13th annual conference on Computer graphics and interactive techniques, New York, NY, USA, pp. 151-160, 1986.
- [5] T. Milliron, R. J. Jensen, R. Barzel, A. Finkelstein, "A framework for geometric warps and deformations," ACM Trans. Graph., vol. 21, no. 1, pp. 20-51, January 2002.
- [6] Z. Kun, J. Huang, J. Snyder, X. Liu, H. Bao, B. Guo, et. el, "Large mesh deformation using the volumetric graph Laplacian," ACM Trans. Graph., pp. 496-503, 2005.
- [7] W. T. Reeves, "Particle Systems a Technique for Modeling a Class of Fuzzy Objects," ACM Trans. Graph., vol. 2, no. 2, pp. 91-108, April 1983.
- [8] T. J. Chung, Computational Fluid Dynamics. The Pitt Building, Trumpington Street, Cambridge, United Kingdom: Press Syndicate of Cambridge University, 2002.
- [9] J. Stam and F. Eugene, "Depicting fire and other gaseous phenomena using diffusion processes," Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, pp. 129-136, 1995.
- [10] K. Polthier and M. Schmies, "Geodesic Flow on Polyhedral Surface," in Data Visualization, Springer Verlag. Proceedings, 1999.
- [11] H. Lee, L. Kim, M. Meyer, and M. Desbrun, "Meshes on Fire," in In EG Workshop on Computer Animation and Simulation, pp. 75-84, 2001.
- [12] Z. Melek and J. Keyser, "Interactive Simulation of Burning Objects," in Pacific Conference on Computer Graphics and Applications, pp. 462-466, 2003.
- [13] R. Malladi, J A Sethian, and B C Vemuri, "Shape modeling with front propagation: a level set approach," in Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 17, pp. 158-175, Feb 1995.
- [14] S. Liu, Q. Liu, T. An, J. Sun, and Q. Peng, "Physically based simulation of thin-shell objects' burning," THE VISUAL COMPUTER, vol. 25(5-7), pp. 687-696, 2009.
- [15] X. P. Institut and X. Provot, "Deformation Constraints in a Mass-Spring Model to Describe Rigid Cloth Behavior," in In Graphics Interface, pp. 147-154, 1996.
- [16] K. L. Gay, L. Ling, and M. Damodaran, "A quasi-steady force model for animating cloth motion," in Proceedings of IFIP International, pp. 357-363, 1993.

# Context-dependent Action Interpretation in Interactive Storytelling Games

Chung-Lun Lu

Department of Computer Science  
National Tsing Hua University  
Hsinchu, Taiwan 30013, R.O.C.  
chunglunlu@gmail.com

Von-Wun Soo

Institute of Information Systems and Applications  
National Tsing Hua University  
Hsinchu, Taiwan 30013, R.O.C.  
soo@cs.nthu.edu.tw

**Abstract**—In this paper, a framework of context-dependent behavior interpretation in interactive storytelling system is proposed. A user can act as one of the role characters in a story to interact with other virtual characters in the system. We implemented two levels of action interpretation: activity and behavior. A Microsoft Kinect sensor is used to acquire and recognize user’s activities in terms of the information of its body joints that will be trained by a pre-learned model. Then, with multiple-context modeling and the recognized activities, a dynamic Bayesian network is adopted to disambiguate user’s behaviors in terms of his intentional and subgoal structure.

*Keywords*-interactive storytelling; behavior interpretation

## I. INTRODUCTION

In interactive storytelling, it’s important for users to play some active roles and interact with the storytelling system. Past researches, for example [1], allow a user to play the role as one of the virtual characters in the cast; the user is projected into the virtual world by mixing the user image into the virtual world on a screen, and then the system interprets the user’s actions based on the user’s utterance and gestures. However, in storytelling, due to the variations of a plot, the same user’s actions can imply different meanings under different contexts. Without proper background contexts, it is difficult to interpret the behavior of a virtual agent due to the ambiguities of the actions. So in this paper, we propose a system that could recognize the user’s activities and interpret their underlying intention using the background context information in the virtual environment. The Microsoft Kinect sensor is used to capture the user’s action images in the real world and so that a 3D virtual character in the virtual world can be controlled and directed by the user via “natural” interactions. Once the activities are captured and recognized and with the aid of multiple context models, a dynamic Bayesian network is used to interpret the user’s action further in terms of the intentional and goal structure of the user which we call it the behavior level of action interpretation.

The remaining of this paper is organized as follows: Section 2 describes some related works of interactions in interactive storytelling games. Section 3 describes the presented scenario. The proposed architecture is described in Section 4, and the detailed context definitions are discussed in Section 5. The method of recognizing user’s activities is described in Section 6 and Section 7 describes the method of

interpreting user’s behavior. Finally, we summarize the presented work in Section 8.

## II. RELATED WORK

In 2001, Charles, Mead, and Cavazza [2] proposed a system that uses an unreasonable concept to interactive with the storytelling system, they called it user intervention. Instead of directly interacting with the virtual characters, users play a role like a god and hide important virtual objects, which the virtual character will look for. The missing of important virtual objects forces the system to re-plan in order for the virtual character to achieve his final goal. In 2004 [1], they improved the way of interaction; a user is allowed to play the role as one of the virtual characters in the cast. The user’s actions are interpreted based on the user’s utterance and gestures; however, the action interpretation in their work didn’t take the variations of contexts into account. In 2010, Doirado and Martinho [3] proposed a system that allows a user to interact with the virtual world and a virtual observer, a virtual dog and detects the user intentions while the user moves; however, in their work, the system achieves the interpretation based on the distance measurement only. In this paper, we aim to interpret the user’s actions based on higher level information, that is, the contexts.

## III. DESCRIPTION OF THE SCENARIO

The scenario of the story presented in this paper is based on a famous Japanese detective comic. Phantom Thief Kaito, a well-known thief, using a fake name, Sot, to work as a butler in the house of Duke Edward. One day, Duke Edward found that one of his antique paintings was missing, so he hired Conan, a brilliant detective, to investigate this crime. After a few days, Conan found that there were three suspects involved, the butler Sot, the female servant Susan, and the guard Alex. The main theme of the story scenario is for the major character, the detective Conan, to investigate this criminal case and find out the actual thief.

Fig. 7 shows an example of the subgoal structure that may be used by the detective. The main goal of the detective is to find out the thief and the detective has two sub-goals, collect evidence information (“Collect Information” goal) and conclude the criminal case (“Conclude Case” goal). To collect sufficient evidence information, the detective needs to interrogate the three suspects (“Interrogate Suspects” goal) and examine the house (“Examine House” goal) to collect

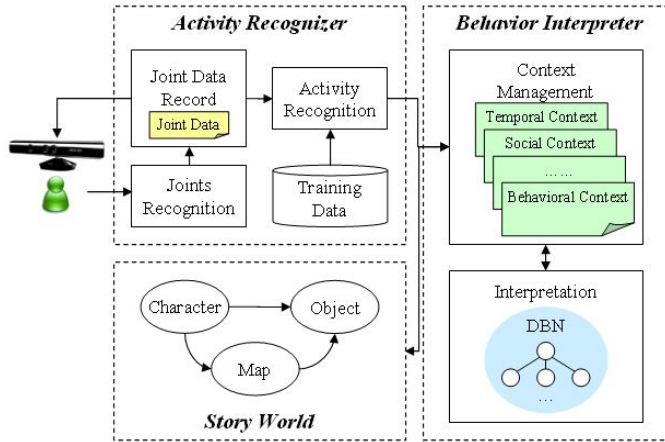


Figure 1. System architecture proposed in the paper.

objective evidence. In this paper, we assume the user play the role of Conan, the detective, and therefore, the user’s activities will be captured via a camera and interpreted by the system.

IV. SYSTEM ARCHITECTURE

The architecture of the proposed interactive storytelling system is shown as Fig. 1. The system consists of three modules: the activity recognizer, the behavior interpreter, and the story world. The functions of each module are described as follows:

- Activity Recognizer

The activity recognizer generates the recognized activity for the behavior interpreter. The user activity is captured via Microsoft Kinect sensor, and then Microsoft Kinect SDK [4] is used to recognize the joints of user’s body. The activity recognizer records the joint data until a pre-defined activity is detected and recognized. The output of this module, the recognized activity, is then sent to the behavior interpreter. In this paper a behavior is defined by an activity with semantic meaning in terms of goal and intention.

- Behavior Interpreter

The behavior interpreter manages the context change according to the story plot and the recognized activities and interprets the recognized activities based on the context information. The detailed interpretation scheme is discussed in Section 6.

- Story World

The story world in this paper is implemented using Unreal Development Kit (UDK) [8]. A virtual world map is implemented that has two rooms, a lobby and a small kitchen in the house of Duke Edward, shown as Fig. 2. The virtual characters are implemented using Autodesk Maya [5] and then imported into Unreal Development Kit, as shown in Fig. 3.

V. CONTEXT DEFINITION

The contexts used in this work are classified into five categories: temporal context, spatial context, social context,

emotional context, and behavioral context. Following is the detailed definition for each type of context.

- *Temporal context* and *spatial context*, including *state* and *location*, evolves according to the storyline. *State* is defined as the node in the hierarchical task network, as in Fig. 7, and *location* is the place where the detective is, a lobby or a small kitchen in the house of Duke Edward.
- *Social context* is defined by *role* and *relation*. *Role* is a set of pairs,  $\{(name, role), \dots\}$ , that denotes each character’s social role and relations in a storyline; for an instance,  $\{(Conan, detective), (Sot, butler)\}$  means the Conan is a detective and Sot is a butler. *Relation* is a three-element set,  $\{(name, name', relation)\}$ , that describes the social relation between characters; for an instance,  $\{(Sot, Alex, colleague)\}$  means Alex is Sot’s colleague.
- *Emotional context* denotes the feeling state of characters; the *feeling* of a virtual character is indicated as a set of pairs,  $\{(name, feeling state), \dots\}$ , and the value of *feeling state* is one of the eight primary emotions proposed by R. Plutchick [7] plus the neutral emotion. The eight primary emotions are joy, sadness, fear, anger, surprise, anticipation, trust, and disgust. In this work, the emotional context is evolved while the system recognizes a user behavior successfully and its value is designed in the background story beforehand.
- *Behavioral context* identifies the activity of a human-played character. This context is composed of recognized activities and its target character or object of this activity,  $(activity, target)$ . Fig. 4 shows an example of contexts.



Figure 2. A scenario scene of virtual world.



Figure 3. Virtual characters are imported into Unreal engine.

**Spatial/Temporal Context:**  
 - *location*: lobby  
 - *time*: evening 9 PM

**Social Context:**  
 - *role*: {(Conan, detective), (Sot, butler), (Susan, Servant), (Alex, guard)}  
 - *relation*: {(Conan, Sot, authority), (Conan, Susan, authority), (Conan, Alex, authority), (Sot, Susan, colleague), (Sot, Alex, colleague), (Susan, Alex, colleague)}

**Emotional Context:**  
 - *feeling*: {(Conan, neutral), (Sot, neutral), (Susan, neutral), (Alex, neutral)}

**Behavior Context:**  
 - *activity*: approaching  
 - *target*: Sot

Figure 4. An instance of a context.

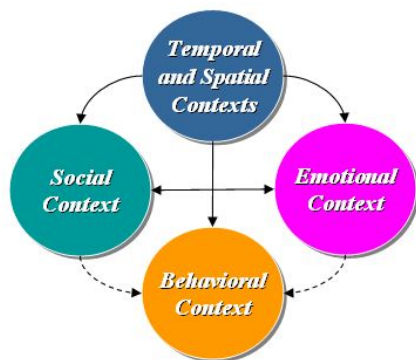


Figure 5. Relation among contexts.

Temporal and spatial contexts may be changed by the storyline and this change may result in different social and emotional contexts; for instance, once the detective knows that Sot is the thief who stole the painting, the main social role of Sot is changed from a butler to a thief. The change of main *social role* may cause *feeling* state to evolve; after the detective figures the crime out, the feeling state of the thief is changed to fear. Behavioral context differs from other contexts; it can be viewed as an independent one since its value is triggered by the user’s activities.

Figure 5 shows the relations between different types of contexts. The changes of temporal or spatial context may trigger the changes of social, behavioral, or emotional context. The social context and the emotional context can influence each other. Under some situations, the behavioral context may be influenced by social or emotional context; however, in this paper, the behavioral context is triggered by the user’s activities, so dotted line is used to represent that the change may not happen in this work.

VI. RECOGNIZING USER ACTIVITY

To recognize the user activity, three steps are necessary. First, body joints of a human need to be captured and recognized. Microsoft Kinect sensor is used to capture the



Figure 6. Result of connecting Microsoft Kinect and UDK.

body joints and NIUI is used to connect Kinect sensor and UDK; NIUI stands for OpenNI/Kinect API for UDK [6]. The result of connecting Kinect and UDK is shown as Fig. 6. We train the recognizer using machine learning techniques, for example, boosting or SVM. Finally, the model is used to recognize the activities and the recognized result is then sent to the behavior interpreter as an input of behavior context.

VII. INTERPRETING USER BEHAVIOR

The same user’s activities can imply different intention under different background contexts. For example, if the detective approaches the butler at the investigating stage, it is more likely the detective wants to interrogate the suspects; if the detective approaches the butler at the case conclusion stage, it means the detective already knows the butler is the thief who stole the painting and he is trying to catch the butler.

Bayesian network has the advantage that each node’s conditional probability distribution can easily be estimated, but the traditional Bayesian network cannot handle temporal data, that is, the new coming data cannot have any contribution to update the model. In interactive storytelling, contexts change over time, so in order to interpret the user behavior at anytime correctly, a dynamic Bayesian network is more appropriate than traditional Bayesian network. To model a dynamic Bayesian network, four parameters need to be defined:

- **Hidden state**,  $Q_t$ , specifies various context states we want to interpret; in our contexts, hidden states are defined by a set of context states. Based on the contexts, the system could interpret the user’s activities in terms of the goal states of the detective defined in Fig. 7.
- **Transition model** is represented by the transition probability distribution functions (pdfs),  $P(Q_t|Q_{t-1}, A_t)$ ;  $Q_t$  and  $A_t$  are the hidden state and action at time  $t$  respectively.
- **Observation model**,  $P(Y_t|Q_t)$ , specifies the dependency of the observation nodes according to the hidden nodes at time  $t$ ;  $Y_t$  is the observations at

time  $t$ . In this work, observations are defined by a subset of contexts; that is,  $Y_t$  is the sensed contexts at time  $t$ .

- **Initial state distribution**,  $P(Q_0)$ , represents the probability distribution in the beginning of the story.

The dynamic Bayesian network is to find the probability distribution of the corresponding hidden states given a set of observations at time  $t$ , as in (1).

$$P(q_t | y_t) = r_t / \sum_{q_t} r_t \tag{1}$$

where  $q_t$  is a set of  $t$  consecutive observations,  $y_t$  is a set of the corresponding hidden states, and the equation of  $r_t$  is in (2).

$$r_t = P(y_t | q_t) \cdot \sum_{q_{t-1}} (P(q_t | q_{t-1}, a_t) \cdot P(q_{t-1})) \tag{2}$$

The user behavior interpretation may face the problem of ambiguity; to reduce the ambiguity, the hidden node  $q_t$  with the highest probabilities in the dynamic Bayesian network is chosen as the most appropriate interpretation.

VIII. CONCLUSION AND FUTURE WORK

In this paper, a framework of context-dependent behavior interpretation in interactive storytelling system is proposed. This paper aims to propose a new way to interact with the storytelling system and adds the context-dependent behavior interpretation into interactive storytelling. Thanks to the release of the Microsoft Kinect, now it's convenient for researchers to implement an interactive system with body movement. Also, dynamic Bayesian network allows the changes of the plots to be triggered by a user who acts as one

character of the cast and can update the probability distribution of states under various context change and thus support the reasoning backward and forward among hidden states, actions and observations along a sequence of activities; it supports the system in interpreting the user behavior at certain accuracy if the model parameter information is collected to some extent. Hence, the proposed system has high feasibility to provide a new interactive experience in interactive storytelling.

ACKNOWLEDGMENT

This research was supported in part by the National Science Council of Taiwan, R.O.C. under Grant NSC 99-2221-E-007-090-MY3.

REFERENCES

- [1] M. Cavazza, F. Charles, S. J. Mead, O. Martin, X. Marichal, and A. Nandi, "Multimodal Acting in Mixed Reality Interactive Storytelling", IEEE Multimedia, July-September 2004, Vol. 11, Issue 3, pp. 30-39.
- [2] F. Charles, S.J. Mead, and M. Cavazza, "User Intervention in Virtual Interactive Storytelling", Virtual Reality International Conference, 2001.
- [3] E. Dourado and C. Martinho, "I Mean It! Detecting User Intentions to Create Believable Behaviour for Virtual Agents in Games", Proceedings of 9th International Conference on Autonomous Agents and Multiagent Systems, 2010.
- [4] Kinect for Windows SDK, <http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/>
- [5] Maya, <http://usa.autodesk.com/maya/>
- [6] OpenNI/Kinect API for UDK (NIU), <http://forums.epicgames.com/showthread.php?t=765636>
- [7] R. Plutchik, "The Nature of Emotions", American Scientist, 2001, pp. 244-251.
- [8] Unreal Development Kit, <http://www.udk.com/>

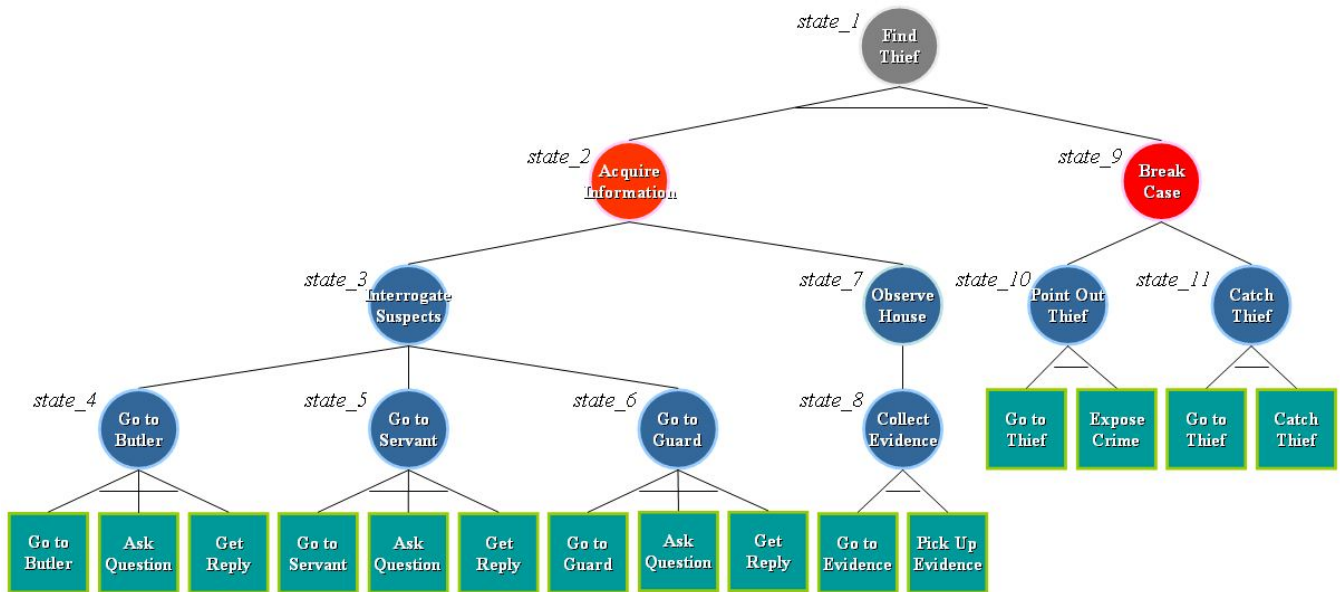


Figure 7. Example scenario of the detective storyline presented in this paper.

## Developing user-centered video game concepts for language learning

Yorick Poels, Jan-Henk Annema, Bieke Zaman  
 Centre for User Experience Research (CUO)  
 K.U.Leuven / IBBT Future Health Department  
 Leuven, Belgium  
 {yorick.poels, janhenk.annema,  
 bieke.zaman}@soc.kuleuven.be

Frederik Cornillie  
 Interdisciplinary Research on Technology, Education &  
 Communication (iTec)  
 K.U.Leuven / IBBT Future Health Department  
 Kortrijk, Belgium  
 frederik.cornillie@kuleuven-kortrijk.be

**Abstract** — This paper will report on an ongoing project which aims to develop video games for language learning through a user-centered and evidence-based approach. Therefore, co-design sessions were held with adolescents between 14 and 16 years old, in order to gain insight into their preferences for educational games for language learning. During these sessions, 11 concepts for video games were developed. We noticed a divide between the concepts for games that were oriented towards formal language learning (e.g. exercises on vocabulary) and video games that were centered around communication with other players or in-game characters.

**Keywords** - co-design, user-centered, games, language learning.

### I. INTRODUCTION

Nowadays, video games are no longer designed solely for entertainment purposes. The resurged interest in serious games shows that many areas can benefit from the engaging experience that video games offer. For instance, video games have been designed to help people in various therapeutic contexts [8, 27], as well as for explicitly educational purposes [21, 22, 26]. In the field of Computer-Assisted Language Learning (CALL), games have been developed specifically for language instruction [12, 13, 20], and have, to a more limited extent and much more recently, been subjected to empirical research [4, 14, 18, 24].

One reason why games may be particularly suited for language learning is that many aspects of video games, for instance problem/puzzle solving and scoring, are also present in language learning [16].

On a broader level, language fulfills a meaningful role in games so that players have to use it as a resource to advance in the game, which resembles the way language is used in functional approaches to language teaching [7, 17]. For instance, in many games language plays an important role in the interaction between the user and the game, e.g. through game scripts and in-game dialogues, and players need to comprehend and/or use language in order to approach goals which are meaningful for them. Also, with the recent rise of massively multiplayer online games (MMOGs), language is a crucial means of communication between players of different linguistic and cultural backgrounds, who need to communicate effectively in order to achieve collaborative goals [25].

It may be argued that precisely because of this apparently natural link between games in which dialogue, narrative and

communication play a significant role and the currently predominant pedagogical focus on functional approaches to language instruction, little research has been devoted to mini-games. Mini-games are small, self-contained games which usually take a short amount of time to complete and which focus on a specific topic [6]. Mini-games are ubiquitous, and have been developed for several purposes, such as for education [6], to offer distraction, [15, 23] and as part of regular video games, for example, small puzzles or simple sports mini-games that are embedded in the Final Fantasy RPG series [23]. In contrast to video games that have traditionally been developed for personal computers and gaming consoles, mini-games can also be found on mobile phones and the web.

Due to their relatively limited size and complexity, mini-games are less expensive, which is an advantage considering the limited budgets of serious games. Furthermore, mini-games can be easily reused in other contexts. This might make them particularly suitable for formal language instruction, especially for explicit vocabulary and grammar training, which often feature short and repetitive exercises. However, little is yet known about how language learners might respond to such games, apart from the fact that learners from a difficult socio-cultural background seem to prefer them over fully immersive games [10].

If it were not challenging enough to design full video games for foreign language instruction that are both effective in terms of language acquisition and at the same time are motivating [17], designing mini-games explicitly for foreign language teaching is probably even more of an ordeal. It is conceivable that from the perspective of a learner/player, certainly if (s)he plays voluntarily, gaming is an end in itself rather than a vehicle to learn a language [11]. Hence, language learning exercises should not merely be disguised as a game, but integrated as part of a game that is fun to play. As a result, it is mainly whether the learner perceives an educational (mini-)game as a game that determines it will be used as such, not what teachers or instructional designers intend to do with it. Primarily, players should want to play the game for pleasure instead of for other reasons.

However, what is needed to create a language learning game that delivers a foremost fun and engaging experience remains unclear. For the design of educational mini-games, this is complicated by the fact that in mini-games the focus is on language as a formal system, which might be less

motivating than when the focus is on language as communication. This underlines the need for a user-centered approach in designing such games.

Therefore, in the study reported in this paper, we set out to examine what it is that makes a game for language learning engaging and something the learners want to play for the sake of fun alone. The study specifically focused on second language learning for adolescents and the needs of these learners concerning fun and engaging educational games for language learning.

*Discovering needs of the learner: co-design*

To design an engaging game experience, it is essential to understand the needs of the people who will play the game in the future. Sanders [19] describes three perspectives from which a person’s experience can be explored.

The first perspective is to listen to what people say and think by using methods from the market research field such as questionnaires, focus groups and interviews. However, by only listening to what people say, important information is missed. People often forget essential details when explaining their day to day business to others and do not know what information is important for the design process [9].

The second perspective is to look at what people do. Different methods from the user-centered design field can be used to observe what people do and use. This way, more insight can be gained into the details of what people do, how they do it, and, for instance, how their environment restricts them in what they do [9].

However, to create a game that is truly engaging, it is necessary to also understand what people know, feel, and what they dream of. This information is hard to uncover as it may not readily be expressed in words, or cannot be observed as it might, for instance, be about latent needs. One way for researchers and designers to get insight into these aspects of experience, Sanders [19] argues, is to have people expressing their thoughts, feelings, and dreams by making artifacts (see Figure 1).

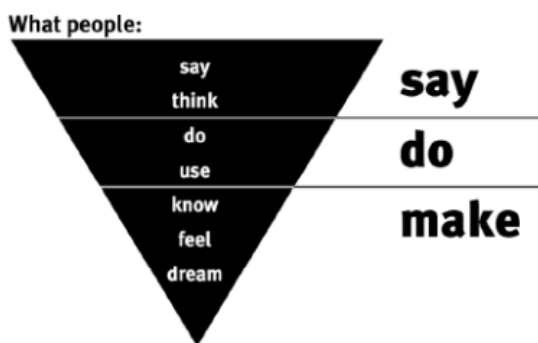


Figure 1. Sanders: “Ways we can learn from people” [19]

One way to do this is through co-design sessions in which users create things together with designers and

researchers. By exploring ideas and concepts together and by making and evaluating artifacts, users communicate directly with the designers and researchers. The artifacts that are made can be considered low fidelity prototypes of the future application. Co-design is a method to gain insight into the participants’ ideas and point of views on certain topics, language learning in games for instance, through the use of a creative process. The results should not be regarded as finished designs. Nonetheless, for social scientists and human computer interaction researchers, co-design can be a real valuable asset in discovering the needs of end-users.

In order to gain as much as possible from co-design it has been found that creating multiple prototypes is more effective than creating a single prototype. A study by Dow et al. [5] showed that when it comes to the design outcomes, exploration, sharing, and group rapport, creating multiple prototypes was the better option.

II. METHOD

A. Participants

A total of fourteen adolescents participated in the study. Two sessions were held. One session in the forenoon, which consisted of eight participants, and one afternoon session, in which 6 adolescents participated. All participants were between 14 and 16 years of age, only one of the fourteen participants was a girl. Twelve of the participants were in general secondary education (called ASO in Belgium), while two participants were from the technical secondary education (TSO). The participants were recruited through online forums, electronic newsletters, paper flyers and posters.

The eight participants of the first session played on average 41 minutes a day. Two of these participants stated that they never played videogames, while four played games more than one hour a day. The six participants of the second session played on average 1 hour 35 minutes. Two of the six participants played three to four hours a day, while the others played more on weekend days.

B. Procedure

Two sessions were held; each session lasted approximately three hours. Each session consisted of an introduction round, a group discussion, a game design round, and a concluding group discussion. This procedure was selected to follow the typical cognitive process of creativity as closely as possible. This process is typically divided into four or five different stages [1, 2, 3].

1) *Introductory round*

Using a slideshow presentation the topic at hand and the co-design methodology were explained briefly to the participants. Then, examples from previous co-design sessions were presented. These examples were taken from domains other than language learning, in order to prevent possible biasing the creative thinking of the participants. The introduction took around 15 minutes.

2) *Group discussion*



After the introduction the participants were split into two groups consisting of three or four participants. Two researchers joined each group and started a short moderated group discussion, in which current language learning practices, both formal and informal, as well as their general experience with learning through games were addressed. This group discussion was intended as a ‘sensitizing activity’, which is a typical first stage in a creative process. This group discussion lasted 20 minutes.

3) *Game design round*

Each group was asked to come up with three game concepts and created low fidelity paper prototypes of these concepts using the available materials (see Figure 2). The participants got one hour to create the three prototypes. The prototypes by no means had to be complete designs.

No explicit instructions were given whether the participants had to create mini games or regular video games. We did not want to limit the participants to only develop mini games, in order to provide them with an opportunity to think freely about games for language learning. However, when a group of participants had come up with two concepts that were clearly regular video games instead of mini-games, the researchers stimulated the participants to make their third design a mini game.

When looking at the different stages of the creative process [19], this round resembled the third stage, inspiration. In this stage, possible solutions or new insights typically occur. The second stage, incubation, was not present in our study due to practical concerns, as the sessions were scheduled on one day.



Figure 2. Creating a video game prototype

4) *Group discussion / interview*

After the game design round, the two groups of participants were asked to present their prototypes to each other and the researchers (see Figure 3). The participants in the other group could ask questions, comment on the prototypes, and had to pick the best design in their opinion. The researchers asked questions to clarify the design choices. After both groups had presented their prototypes, a group discussion was started that revolved around a number of pre-defined topics such as user-oriented and personal goals within the

game, the role of the teacher, context and the use of the video game.

The group discussion can be seen as an activity that represents the final stage, transformation, in a creative process [2]. An evaluation takes place during the transformation stage, to decide whether an idea is valuable. The ideas are elaborated upon to validate and communicate them with the rest of the group. This discussion took about an hour.



Figure 3. Prototype presentation and group discussion

III. RESULTS

The created game concepts were varied and demonstrated a range of elaborateness. Some concepts were original, whilst other concepts adhered more to generally known game concepts or game genres. In general, two major categories could be discerned. On the one hand, games that were focused on formal learning, the principle of language learning with, for example, vocabulary exercises. On the other hand, game concepts were created that used language as a means of communication. In the following two sections, we will give an overview of these two categories.

A. *Games for formal language learning*

Three of the game designs that were created, were aimed towards formal language learning. These games shared the following characteristics: focus on vocabulary, immediate feedback, limited time requirement, and little or no narrative.

These characteristics will be outlined based on one of the game prototypes that was developed, the cannon-versus-monsters game (see Figure 4). In this game, the player has to translate a word as quick as possible in order to prevent monsters, descending on a narrow path, to reach the player. The number of bullets a player receives depends on the length of the assigned word. For instance, a four letter word that is correctly translated gives the player four bullets to eliminate the approaching monsters (see Figure 4). The difficulty level gradually rises with each stage, offering the player more challenging words to translate, but also offering

more bullets and useful power-ups (any item that temporarily gives a character new abilities, new powers, or a statistical bonus).

As said above, three of the game concepts focused on vocabulary. In the cannon-versus-monsters game, for instance, the goal is to survive as long as possible by correctly translating words. In the other two game concepts, vocabulary also is the main language learning topic, as players have to translate words, recognize words or associate them with other words.

Feedback in the cannons-versus-monsters game is provided immediately to the player. Every time the player fails to translate a word, the monsters come closer to the player's home, eventually destroying it when the monsters come near enough. Thus, it is obvious that it is crucial that as many words as possible are translated correctly. When the player fails to translate a word, the consequences are instantly visible as the monsters further approach the onscreen character of the player. In the other two game concepts focusing on vocabulary, players also instantly lose a live or turn when making a mistake.

The cannon-versus-monster game revolves around the relatively simple goal of keeping the monster away. By translating words correctly ammunition is earned that can be used to shoot the monsters. No further narrative or plot was provided as context for the game. This characteristic was shared by the other two game concepts that focused on vocabulary. These games had a very limited, if not, non-existent story line.

Finally, as the cannon-versus-monsters game concept and the other two game concepts focusing on vocabulary were relatively simple and short games, they did not require a lot of time to complete. Thus, these games could easily be played in situations where little time is available. Participants commented in the group discussion that because of this, the games could be used at school as part of language learning classes.

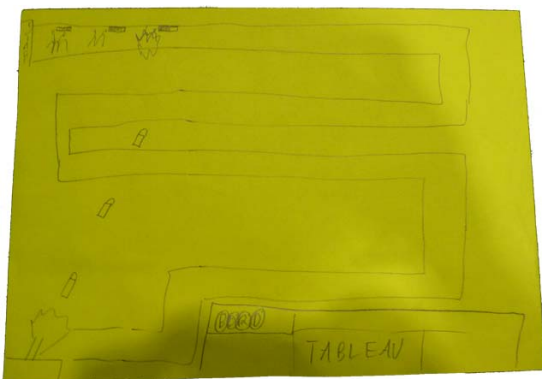


Figure 4. Drawing of the cannon-versus-monsters game concept

*B. Games and language as communication*

Six of the games concepts could be characterized by their focus on language as a means to communicate in the game.

Players needed to communicate to progress through the game. When compared to the three game concepts that focused on vocabulary, these concepts were more complex and contained an elaborate narrative, provided less immediate feedback, and were more time consuming. These characteristics will be discussed by looking at one of these six games concepts, an adventure game, in more detail.

In this adventure game, the player has to get from Paris, France, to Los Angeles, USA, to visit his sick mother. To achieve this, the player has to communicate with other game characters or other non-player characters. Thus, language is the means to get to the end goal. Through dialogues and creative use of language (e.g. asking for a lift, lure opponents into traps, persuasion, deceiving, ...), the game character progresses through the game.

The game concepts that focus on communication were more elaborate than the game concepts that focused on vocabulary. While the vocabulary game concepts contained no or very little narrative, whereas the narrative was very important, and much richer, in the games focusing on communication.

A consequence of the focus on communication and the related, more elaborate, narrative is that the feedback is less immediate than in the games that focus on vocabulary. While in the cannon-versus-monsters game, the player immediately receives bullets to keep the monsters away, or sees the monsters approaching further after each mistake, the progress in the games focusing on communication is less immediately visible. Although the end goal is clear, the player only slowly approaches this goal; the rewards lie on a higher level.

Finally, compared to the games focusing on vocabulary, the game concepts with a focus on communication are relatively complex and therefore required more time to play. The player has to have a period of uninterrupted time available to play. This would make these games, according to our participants in the group discussion, more suited for playing at home.

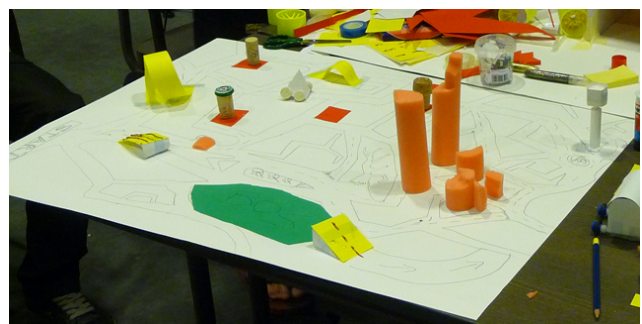


Figure 5. Free roaming game

*C. General characteristics*

Based on the game concept prototypes, we could discern a number of general characteristics for the games developed by our participants. Four of the eleven games had a multiplayer mode. In a number of games, this feature was

not explicitly mentioned, and was open for interpretation. Six of the games incorporated a social component, like the ability to share high scores with friends, and communicate via voice chat.

The choice of the platform (computer, console, mobile) was not specified for most game concepts. Some games were thought to be more suited for a specific platform than others, with game concepts ranging from a traditional mini-game on a desktop computer, to an augmented reality game on a mobile phone.

Concerning reward mechanisms, different concepts were developed by the participants. This ranged from simple scoring systems, e.g. traditional high scores, to more complex rewarding mechanisms, where the player could gain experience points on different levels.

#### *D. Context – instruction – in game feedback*

The context where one could play a game differed also between the various games. Three games were found to be more suitable by the participants for home and school usage. One game they would rather like to play in a home context. For the majority of game concepts, the participants did not define the ideal playing context.

The participants also indicated that in-game feedback mechanisms were of certain importance. They found it important that there is a kind of feedback present in games, for example when the player is stuck. Feedback mechanisms varied from a built-in translator to an in-game character that aided the player as an interpreter for foreign languages.

#### IV. DISCUSSION

This study indicated that there is a divide in games for language learning. On the one hand, our participants developed games for formal language learning that resembled the definition of a mini-game [6]. On the other hand, we discerned games that focused on language as communication. Nevertheless, we were surprised to see that none of the games focusing on formal language learning dealt with grammar.

As we already pointed out, some games were more focused on learning vocabulary, formal learning, while the other games were based around communication. This finding is not necessarily a disadvantage for the design of educational games for language learning, but confirms that games could be used as a medium to create a need for the language learner to accomplish objectives that lie outside the language itself. We could use the game concepts that were developed to identify these needs and incorporate them in an additional layer around the games.

Our study posed a number of limitations. First, there was the sample that consisted mostly of adolescents of the same school level (general secondary education). As our study had no intention to generalize these results to a larger population (all school levels), the limited sample did not pose a threat for the explorative nature of our study. The participants were mainly male adolescents. We suspect that they were more prone to react to our message. We did, however, spread the recruitment message via numerous channels, online as well

as offline (posters and flyers in schools, public library) in order to prevent overrepresentation of certain groups (hardcore gamers, tech savvy boys). However, there still was more reaction from adolescents from the ASO, and also more from boys.

Second, we noticed that the participants created games or game concepts that, sometimes, resembled existing games. This could point in different directions, namely that our participants had played a lot of commercial titles and were influenced by these concepts, or that they had little imagination and therefore copied existing games. Future research could link the participants' gaming history and preexisting preferences to the creation of game concepts in co-design sessions.

And finally, from a game design point of view, co-design is not an ideal method for developing fully finished games. Co-design is a method for researchers to gain insight into the participants' needs and experiences concerning language learning in games. Nonetheless, for social scientists or human computer interaction researcher, co-design can be a valuable asset in discovering the needs of end-users, through a creative process.

#### V. CONCLUSION

In sum, the results of this co-design study revealed insights into game concepts for the design of video games for language learning, from a user centered design perspective. These insights can be interesting for the development of language learning games.

We noticed a divide between the concepts for mini-games, that were oriented towards formal language learning (e.g. exercises on vocabulary) and video games that were based around communication with others (players or in-game characters).

#### ACKNOWLEDGMENT

This research was conducted as part of the IBBT-MiGaMe project. The work was supported by K.U.Leuven and funded by the IBBT (Interdisciplinary institute for BroadBand Technology), a research institute founded by the Flemish Government in 2004.

REFERENCES

1. Boden, M.A. *Creative Mind: Myths and Mechanisms*. Routledge, New York, NY, 10001, 2003.
2. Bullinger, H.-J., Müller-Spahn, F., and Rössler, A. Encouraging Creativity - Support of Mental Processes by Virtual Experience. *Virtual Reality World*, (1996).
3. Csikszentmihalyi, M. *Creativity: Flow and the Psychology of Discovery and Invention*. Harper Perennial, 1997.
4. deHaan, J., Reed, W.M., and Kuwada, K. *The effect of interactivity with a music video game on second language vocabulary recall.(Report): An article from: Language, Learning & Technology*. University of Hawaii, National Foreign Language Resource Center, 2010.
5. Dow, S., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D., and Klemmer, S. Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. *Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM (2011), 2807-2816.
6. Frazer, A., Argles, D., and Wills, G. Assessing The Usefulness Of Mini-games As Educational Resources. *ALT-C 2007: Beyond Control*, (2007).
7. García-Carbonell, A., Rising, B., Montero, B., and Watts, F. Simulation/gaming and the acquisition of communicative competence in another language. *Simul. Gaming* 32, 4 (2001), 481-491.
8. Griffiths, M. The therapeutic use of videogames in childhood and adolescence. *Clinical child psychology and psychiatry* 8, 4 (2003), 547.
9. Hackos, J.A. and Redish, J.C. *User and Task Analysis for Interface Design*. John Wiley & Sons, Inc., 1998.
10. Herselman, M.E. The application of educational computer games in English second language teaching. 1999.
11. Hubbard, P. Evaluating Computer Games for Language Learning. *Simulation & Gaming* 22, 2 (1991), 220 -223.
12. Hubbard, P. Interactive Participatory Dramas for Language Learning. *Simulation & Gaming* 33, 2 (2002), 210 -216.
13. Johnson, W.L. Serious Use of a Serious Game for Language Learning. *Proceeding of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, IOS Press (2007), 67-74.
14. Miller, M. and Hegelheimer, V. The SIMs meet ESL. Incorporating authentic computer simulation games into the language classroom. *Interactive Technology and Smart Education* 3, 4 (2006), 311-328.
15. Miniclip. Miniclip Games - Play Free Games. 2011. <http://www.miniclip.com/games>.
16. Prensky, M. Digital game-based learning. *Comput. Entertain.* 1, 1 (2003), 21-21.
17. Purushotma, R., Thorne, S.L., and Wheatley, J. 10 Key Principles for Designing Video Games for Foreign Language Learning - een knol van Ravi Purushotma. 2008. <http://knol.google.com/k/10-key-principles-for-designing-video-games-for-foreign-language-learning#>.
18. Ranalli, J. Learning English with The Sims: exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning* 21, 5 (2008), 441-455.
19. Sanders, E.B.N. From user-centered to participatory design approaches. *Design and the social sciences: making connections* 2, (2002), 1.
20. Sanders, R.H. and Sanders, A.F. History of an AI spy game: Spion. In *Thirty years of computer assisted language instruction*. CALICO, 1995, 114-127.
21. Serious Games Interactive. *Global Conflicts: Palestine*. Gamers Gate, Manifesto Games & Macgamestore, 2007.
22. Spongelab Interactive. *Genomics Digital Lab - History of biology*. Spongelab Interactive, 2009.
23. Square Enix. *FINAL FANTASY*. 2011.
24. Sykes, J.M. A dynamic approach to social interaction: Synthetic immersive environments & Spanish pragmatics. 2008. <http://purl.umn.edu/91603>.
25. Thorne, S.L., Black, R.W., and Sykes, J.M. Second Language Use, Socialization, and Learning in Internet Interest Communities and Online Gaming. *The Modern Language Journal* 93, (2009), 802-821.
26. United Nations World Food Programme. *Food Force: The First Humanitarian Video Game*. World Food Programme, 2005.
27. Vanden Abeele, V., Geurts, L., Husson, J., Windey, F., Annema, J., Verstraete, M. & Desmet, S. Designing Slow Fun! Physical Therapy Games to Remedy the Negative Consequences of Spasticity. *Proceedings of the 3rd International Conference on Fun and Games*, ACM Press (2010).

# Online Casinos: The Addiction Under Control

*Role of Web 2.0 and re-documentarisation.*

Karim Fraoua, Christian Bourret

Equipe Dispositifs d'Information et de Communication à l'Ère Numérique (DICEN)

Université Paris-Est

77454 Marne-la-Vallée, France

fraoua@univ-mlv.fr, bourret@univ-mlv.fr

**Abstract**—In this work, we will consider how to design and implement tools or Information System based and enriched by a semantic search engines to the forums in the world of online gambling (articles, tutorials, etc.) supporting both textual and semantic expressions. On the basis of this new approach, based on neuro-economic field, we will look on addiction using neuroscience and game theory in order to construct an efficient Information System that takes into account the expectations and attitudes of players.

**Keywords**—online; gambling; addiction; neuroeconomy; game theory; information system.

## I. INTRODUCTION

For a long time the game became a regular and acceptable part of our society. For most people, gambling is a social activity that is fun and entertaining. Besides the casino, companies insist on this aspect to indicate that this is primarily a leisure activity. But for others, gambling problem cause a real addiction phenomenon similar to drugs: they become uncontrollable and the game is no longer a choice for them [1]. Studies on the game take into account gambling behavior including those conducted by the casino companies. Recently a study of gambling problem in France shows clearly that this problem is growing, with more than 600 000 players may be considered as presenting a risk of addiction or considered as addict [2].

The online game is a new element in our modern society, creating a new type of players and new audiences. The challenge today is to consider the broad issue of addiction to online gamers and bypass techniques possible to ban it. It is now quite possible to manage the technical aspect of changing IP address or credit card during elicitation for gambling anonymous IP addresses to be managed by operators. The issue of access of online gaming raises new questions that are less marked for classical casino.

We will see in this paper that we can bring in new sight on addiction problem using neuroscience approach and how to design a new scheme in re-documentarisation of forums that permit to detect an addictive player and to orient him to a medical space, where his problem can be treated.

## II. THE ADDICTION PHENOMENON

The addiction to gambling has always existed. It has been described by writers such as Dostoïevski in “The Player” [3]. This problem affects the mental health of some players [4]. It has been studied in some U.S. cities that focus on spaces for whose playing become an important issue for residents [5]. Gambling problems often co-occur with mental health problems, chemical dependency with drug use and of course this affects the family, generating legal and financial problems. This can lead to suicidal behavior with acting out often observed especially when the player has lost all hope of settling the quagmire in which he began by referring to the phenomenon mentioned above.

The behavior of these so-called addictive games players was the subject of several sociological studies, both in behavioral terms [6] and from few years the contribution of neuroscience [7]. Several studies on psychological factors in a rational point of view of addiction exist in the literature [8].

The question of how to handle this problem is essential, we believe that to understand the path of decision, we should look to the issue in terms of neuroscience or neuro-economics, not only in psychological terms.

Thus, the problem arises from the classification of the addiction. Look so far where are focused the addiction, our society consider that it is mainly clinical, although studied in a particular context. It should be noted that the factor game can also be cultural, social or ethnic like in Chinese or Jewish communities [9]. However no study shows the prevalence of gambling addiction except in a population of a certain socio-economic with a lowest wages where research shows sometimes the extreme behaviors, leading to games to match with words addictive behavior.

The reasons why people play to game and why they can becomes addict in general are poorly identified. They can be numerous and varied [10]. Several ideas were discussed, as a social acceptance within a group, the enrichment, a painful event in the life that lead to emotions such as anger, depression, anxiety attacks [11], the desire to test his luck in the game and the excitement of such gaming experience motivated by reward and highlighting the role of dopamine [12].

In general, it seems that the main reason that pushes the people to play, is the desire to win and the denial of loss. The gain is considered as a success, corresponding to the development of self-esteem and recognition by others [13]. To encourage the people to play, the barrier of the loss is now filled by some online operators with an offer of free play at first to lead the player into the game space [14].

Presumably some games are popular because they are based on socialization, real or virtual like for the poker free online website Zynga [15]. In addition, poker is largely diffused in the broadcast on several television stations, and it suffers a major craze, leading to highly addictive attitudes. The addiction to poker is very specific. It can be considered as dangerous because the game can be considered by the players as a rational game; the success depends therefore from the capacity of players and less from chance.

### III. NEW GAME MODALITIES FOR NEW ONLINE PLAYERS

The spectacular development of Internet and mobile telephony has accompanied the new forms of online gambling. Many video games are evolving in the "network" space. The MMORPG (Massive Multiplayer Online Role Playing Game) a networking player system, with significant numbers of players, can now collect up to several million participants. For most of them, these games usually require expensive subscriptions. Mobile telephony also facilitates the appearance of no new quizzes via SMS.

This type of contest mainly aimed to the youngest, offering prize money, new forms of gambling and other subscriptions and downloads dramatically increase the risk of negative consequences (costs, isolation, disinvestment of daily living) for the affected population: adolescents and young adults and large consumers of new technologies [16].

The family aspect can play an important role in the emergence, development and maintenance of the pathology of the game [17]. It is the same on the social learning model, acceptance of the game at the social level, leading the player to understand the game as a normal social activity as it is approved by the legal system.

The attitude of the player as an individual is marked by the traditional social behavior through learning, modeling and maintenance of socially acceptable behaviors and attitudes. The family can be the basis of the driving game, since children often play cards with the family to indicate its acceptance as a normal social activity or when they encouraged to play with other kids [17].

The genetic factors were studied and showing the existence of such a phenomenon of addiction situations with twins. A detailed study of identical twins living together or separated was conducted on 155 twins [18]. However, this study leaves out some aspects that can influence the behavior of players due to the environment of the players or their personalities.

The sociological aspect is the overriding factor. Indeed, in the social model the player tends to enhance the game, because it is less complex and could bring a reward, social and financial value in the gain.

A player will have a form of recognition through his player activity by operators in a form of marketing strategy.

This sociological aspect of escape from the complexity of the social model contributes significantly to the pathology of the game [19].

There is now a neuroscientific approach to the phenomenon of addiction, which encourages new explanations. Indeed, several authors have identified this approach that appears to be an interesting way for understanding the phenomenon of addiction [20].

### IV. NEUROSCIENCE, GAME THEORY AND DECISION CHOICE

This approach can help to renew the study of the issue of addiction. First neuroscience has made considerable progress with brain imaging. Until now the traditional approach treated human behaviors by game theory of Von Neumann and Morgenstern [21] revisited by J. Nash [22].

Indeed, players can easily understand that they are in a situation of asymmetric information and a non-cooperative game. However the player is master of his choice which tends to maximize its usefulness in the game. This utility is either an economic or the well-being.

The second category does not fall within the scope of this investigation. This is a customer, also sought by the operators of casinos coming to play with the primary objective of leisure. The first type of clientele would be subject to the risk of addiction because the main objective is the gain factor or optimization of utility. It is so ironic that this customer led to addictive client because the loss does not seem to be a hindrance to the game

Why do players become addictive even though they are so inconsiderate? In general, in the economic model of the individual player, the choices are for most of them rational. The concept of rationality was originally developed by Von Neumann and Morgenstern.

This first model was useful to explain the economic and social situation at some point although the model assumes a perfect environment result from the theory of general equilibrium. This environment requires a perfect symmetry of information system and cooperation between agents.

Indeed, in economy of goods or services, or other areas in perfect competition, the equilibrium is defined by an economic agent who chooses a solution which does not damage that of another agent. This also requires an atomistic market namely n-players market.

We thus see that the situation resembles to the game in which the agent makes a choice that necessarily depends on the choices of others agents but for which he has perfect information.

The issue of well being or welfare provided by this act depends on the decision taken by other actors. The situation in the middle of the game appears in a model where sharing information is asymmetric. The operator has information that the player does not control, yet it should discriminate on the game where the player participates.

Indeed, although the overall rate of redistribution is known by the players, and it is not known at certain time during the game, whatever the game, the payout of the slot machine in question or another game where player hopes in

a prospective approach to optimize his utility function by building more.

From a mathematical point of view, for a player to be in a position to optimize his choice, he should know all the players participating in the game, the rules of the game and the information available to the other players, which may seem useful in a poker tournament for example, and individual preferences.

When the player enters in a physical place, some of these elements will be respected.

For online game, little is known, since the choices are totally made in the total absence of known parameters. This creates a climate of uncertainty, which puts the player in a prime position where he accepts himself the lottery concept.

When the outcome is certain, the choice may lead in our sense to an addictive situation, harmful in economic terms. The action is not related to a gain or loss of money, it is also the case when the player participates in a recreational activity.

The utility theory gives a first explanation for this type of attitude in an uncertain environment as the factor of utility is based on the calculation of hope, that ignoring the decision himself and especially his attitude versus the risk.

It is recognized that most agents are risk averse from the work of Allais [23] and Ellsberg [24]. The question of the decision based on the expected value of utilities has already been asked in the past and it is clear that the expected value is justified in the long run if the game is repeated several times.

Von Neuman and Morgenstern make a response stating that the election is made by considering the value of each share and this is the obvious choice as risk aversion, it is still necessary that each set leads to a result independently.

We can imagine that the addict player is not risk-averse or ignore all of the parameters of the game in its mathematical form.

Risk is defined by a given situation as the objective probability corresponding to a lottery known result, like the game of roulette at the casino or a roll of the dices that the area is classified as uncertain when the outcome is uncertain as game of poker, slot machine, the result of a "football game". John Nash provides an answer to this question by stating that the players do not cooperate, through the example of prisoner's dilemma because they may be placed in information asymmetry. This new approach allows then to explain some forms of choice that we can operate which are optimal in terms of choice or expected utility because we can be devoid of useful information. Lack of resources (knowledge, understanding, etc.) or time does not collect useful information and, consequently, the utility bill is not at its maximum, but at its best. We can have also a behavior leading to a losing attitude or utility or sometimes to a negative attitude

The neuroscientific aspect has been also clearly demonstrated by several works, as seen before. Van Holst et al. showed that brain areas are activated specifically in the case of gambling, that can be said at risk of addiction specially, the ventral tegmental-orbito frontal cortex [25].

## V. VARIOUS TYPES OF PLAYERS

The world of gamers is not an uniform medium. To build an efficient Information system, it is useful to study the general typology of the players. We must not ignore that there are still subtypes namely women where there is evidence that online gambling is like visiting a shopping site, namely a place of socialization [26]. We must consider the category of young people who remain vulnerable to this type of offer [27].

We see that the sociology of players is widely diffused and cannot conceive the intermediation via a general purpose tool, but an "agile" tool that can adapted to the type of players according to the subcategory.

The player is usually acting individually thinking that he can maximize his welfare function, based on beliefs in giving assumptions about the state of the world and the potential usefulness of each of the possible consequences. This is what he hopes calculate but beliefs and superstitions are subjective in nature.

All these questions can be studied nowadays in a new way that takes into account the cognitive approach rather than collective societal behavioral approaches such as fads, but on an individual basis.

This new sight helps to understand the concept of how decision may be irrational, and especially to understand the mechanisms that push players to act this way. In opposite, the rationalist approach indicates that difference in beliefs between the players can be explained solely by differences in their level of information

## VI. IMPLEMENTATION OF AN ONLINE INFORMATION SYSTEM

Training and education seem also necessary that the implementation program for information on how to react and to prevent gambling problems.

In the case of online games, we will need to act via the web interface since the player is isolated from the place of game where all the aspects we have mentioned exist. The existence of the semantic web would be useful to the extent that we could reconstruct the behavior of the player participation in forums which would be included in the online game site where registration is required.

Indeed, the player was registered by duty on the gaming physical place; it will be thus invited to participate in the forum. It may be encouraged by forms of rewards for participation in this forum. For example, they can be invited through a scoring rate of participation, or the development of voucher on the co-branding, *i.e.*, Drinking offers.

On the other hand, for the development of Information system (IS), the main difficulty in creating semantic operators is linked to the fact that the corpus will not be considered as mere undifferentiated text books because their constitution involves a number of criteria of homogeneity as well as contrast on the quality of the ontology.

It will be assumed to constitute a representative set of productions, where each community could develop its own ontology, is difficult. Indeed, this may cause the constitution of ontology with a low level of use; it is the risk of the

purely social approach, since players can develop specific ontology that is very specific to their own point of view of the world of games, *i.e.*, “*j’ai pris une gamelle*” where in French “*gamelle*” may mean “*loosing term*” or a *plate*.

In order to enhance the usability of this ontology, we will look to the constitution of this type of little formal ontology to the status of the data as well as the place of the expert in the reconstruction of corpus via a more developed ontology, which will be designed in this IS. This will define the role of tools in extraction of terms.

We will also propose the practices and the uses to the player that feed the Information System by analyzing in depth their methods and practices research and selection of documents, and their methods of treatment of these documents to insert them in the Information System

The goal is to reach an automated process of alert following automatic extraction of knowledge from texts related to the game. This is in order to automatically extract knowledge concerning the specific terms of the addiction from texts, with a derived method of automatic processing of language similarly to other domain [28].

This will define a tool which would broadcast information to the owners of the sites of online games on the model of information systems for the detection of the signal and the management of knowledge in addiction.

When addictive behavior is detected, a quiz may be sent to the player which will measure the risk of the player to addictive behavior and especially in order to better take in charge this risk and to maintain the spirit of leisure.

The objective targeted in this work is purely informational, because it is the basis of the correction of these phenomena in our connected society that increasingly appears and which primarily affect the most disadvantaged social populations (unemployed, resigned, etc.)

We need to bring elements of knowledge to the operator to provide the finest and most complete information of very low granularity to the players who are no longer in a position to make or to choose among a lot of decisions possibilities to fit their best utility factor.

We can be optimistic if we base our approach on the result of Polemarchakis and Geanakoplos [28] that players will eventually find consensus between them indicating this fundamental result: agents however fails necessary the same consensus than if they had directly communicated all their information. The main result is that this situation does not continue always. Indeed they indicate this major result that the agents can’t disagree forever.

#### *A Re-documentarisation for forum in online casino website*

Usually the specialists organize the document according to a classical model, a summary, an index and eventually put it in certain categories. In a certain way, when the work is made by a specialist and the writer of the document is a professional, the task can be easily made. However, the problem of forum is that the language is likelihood different and the automatic extraction of a semantic scheme or metadata can be slightly different for the attendance. This may causes a very weakest information level finally and is

not exploitable by the specialist of game addiction in our case.

The re-documentarisation can appear in this situation as a good approach to correct these biases [30]. Indeed, the methodology base is to reconfigure the document with creating annotation or new metadata in order to put the document in a right category.

Zacklad et al. [31, 32] has shown that participatory spaces and forums that fall within the socio-semantic web are likely to structure the documentation process associated with collective governance issues. This may be seen as a key to create an enrichment tool to help specialists to better assist the players against the risk of addiction.

#### CONCLUSION

This work is a prospective paper under progress and is related to the appearance of the legal online game in France that creates today important phenomenon of addiction.

Parliament proposes control measures but this remains inefficient despite this goodwill.

It is clear that this control will not be exercised if the operators themselves don’t put in place the online forums with a continuous and automated process that permit the analysis to detect the so-called players with addictive behaviors.

We are looking forward how to implement to re-documentarisation tool in a casino website space. We are also looking for the best ontology that leads to a better guidance in direction of the players, towards specialized sites or to custom monitoring in order to eliminate addictive behavior.

#### References

- [1] M. Bronnec, B. Rocher, G. Bouju and J.L. Venisse, « Jeu et addiction », *Annales Médico-psychologiques, revue psychiatrique*, Volume 168, issue 7, pp. 509-512, September 2010.
- [2] <http://addiction-au-jeu-600-000-francais-menaces-6709456.html>, September 2011.
- [3] The Player, Fiódor Dostoiévsky
- [4] P. Sacco, R.M. Cunningham-Williams, E. Ostmann and E.L. Spitznagel Jr., The association between gambling pathology and personality disorders, Original Research Article, *Journal of Psychiatric Research*, Volume 42, issue 13, pp. 1122-1130, October 2008.
- [5] J. S. Lemmens, P. M. Valkenburg and J. Peter, Psychosocial causes and consequences of pathological gaming, *Computers in Human Behavior*, Volume 27, issue 1, pp. 144-152, January 2011.
- [6] E. Boyle, T. M. Connolly and T. Hainey, The Role of Psychology in Understanding the Impact of Computer Games, *Entertainment Computing*, Volume 2, issue 2, pp. 69-74, 2011.
- [7] E. Duven, K.W. Müller and K. Wölfling, Internet and computer game addiction - a review of current neuroscientific research, *European Psychiatry*, Volume 26, Supplement 1, p. 416, 2011.
- [8] N. Raylu and T. Po Oei , Pathological gambling, a comprehensive review, *Clinical Psychology Review*, Volume 22, pp. 1009-1061, 2002.
- [9] N. Raylu and T. Po Oei , Role of culture in gambling and problem gambling *Clinical Psychology Review*, Volume 23, issue 8, pp. 1087-1114, January 2004.
- [10] G.J. Smith, *Gambling Studies Encyclopedia of Social Measurement*, pp. 75-84, 2005.



- [11] A. Verdejo-García and A. Bechara, A somatic marker theory of addiction, *Neuropharmacology*, Volume 56, Supplement 1, pp. 48-62, 2009
- [12] I. H.A. Franken, J. Booij and W. van den Brink, The role of dopamine in human addiction: From reward to motivated attention, *European Journal of Pharmacology*, Volume 526, issues 1-3, pp. 199-206, 5 December 2005.
- [13] B. Ayin and S. Volkan San, *Procedia-Social and Behavioral Sciences*, Volume 15, pp. 3500-3505, 2011.
- [14] D.H. Shin and Y.J. Shin, Why do people play social network games?, *Computers in Human Behavior*, Volume 27, issue 2, pp. 852-861, March 2011.
- [15] <http://company.zynga.com/games/featured-games>, December 2011
- [16] S. Hwa Hsu, M.-H. Wen and M.-C. Wu, Exploring user experiences as predictors of MMORPG addiction, *Computers & Education*, Volume 53, issue 3, pp. 990-999, November 2009.
- [17] M.E. Gifford-Smith and C.A. Brownell, Childhood peer relationships: social acceptance, friendships, and peer networks, *Journal of School Psychology*, Volume 41, issue 4, pp. 235-228, July-August 2003.
- [18] *Addiction, A disorder of choice*, Gene M. Heyman, Harvard University Press, 2009.
- [19] H. Rachlin, *Economic Concepts in the Behavioral Study of Addiction?* Chapter 5, *Choice, Behavioural Economics and Addiction*, pp. 129-153, 2003.
- [20] T. Takahashi, A neuroeconomic theory of bidirectional synaptic plasticity and addiction, *Medical Hypotheses*, Volume 75, issue 4, pp. 356-358, October 2010.
- [21] J. von Neumann, and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton, NJ, Princeton University Press, 1944, second ed. 1947, third ed. 1953.
- [22] J. F. Nash, *Proc. Natl. Acad. Sci. USA* 36, pp. 48-49, 1950.
- [23] M. Allais, Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'Ecole américaine, *Econometrica*, 21, pp. 503-546, 1953.
- [24] D. Ellsberg, Risk, Ambiguity, and the Savage Axioms, *The Quarterly Journal of Economics*, pp. 643-669, 1961.
- [25] R.J. van Holst, W. van den Brink, D.J. Veltman and A.E. Goudriaan, Why gamblers fail to win: a review of cognitive and neuroimaging findings in pathological gambling, *Neurosci Biobehav Rev.* Jan Volume 34(1), pp. 87-107, 2010.
- [26] T.O. Afifi, B.J. Cox, P.J. Martens, J. Sareen and M.W. Enns, Demographic and social variables associated with problem gambling among men and women in Canada *Psychiatry Research*, Volume 178, issue 2, pp. 395-400, 30 July 2010.
- [27] L. Blinn-Pike, S. Lokken Worthy and J. N. Jonkman, Adolescent Gambling: A Review of an Emerging Field of Research *Journal of Adolescent Health*, Volume 47, issue 3, pp. 223-236, September 2010.
- [28] H. Déjean, E. Gaussier, J.-M. Renders and F. Sadat, Automatic processing of multilingual medical terminology: applications to thesaurus enrichment and cross-language information retrieval, *Artificial Intelligence in Medicine*, Volume 33, issue 2, pp. 111-124, February 2005.
- [29] J. Geanakoplos and H. Polemarchakis, We Can't Disagree Forever, *Journal of Economic Theory*, Volume 28(1), pp. 192-200, 1982.
- [30] J.-M. Salaün, Redocumentarization, Challenge For Information Studies, *Études de communication*, 30, pp. 13-23, 2007.
- [31] M. Zacklad, Documentarisation processes in Documents for Action (DoFA): the status of annotations and associated cooperation technologies. *Computer Supported Cooperative Work*, June, vol. 15, n° 2-3, pp. 205-228, 2006.
- [32] M. Zacklad, Participative documentary spaces and governance. *International Journal of Sustainable Development*, vol. 11, Nos. 2/3/4, pp. 247-261, 2008.

## *Evaluation of User Interface Satisfaction of Mobile Maps for Touch Screen Interfaces*

Ya-Li Lin

Department of Statistics  
Tunghai University  
Taichung, Taiwan  
e-mail: yllin@thu.edu.tw

**Abstract**—User interfaces of mobile maps on mobile and tablet devices with a touch screen interfaces is evaluated in this study. A four-way factorial experiment compares the user interface satisfaction for two mobile touch devices (iPad/iPhone), two map types (Electronic Map/Mixed Map), two keyword searches (Landmark/District), and Chinese input methods (Hand-written/Phonetic notation). The experiment used forty-eight participants, each of whom was assigned three types of environmental spatial tasks: find targets, identify cardinal directions, and identify approximate distances. The individual differences between gender, the possession of a sense of direction and route knowledge on user’s satisfaction were also examined. The results of the study verify that both the reliability and confirmative factor analysis model of the questionnaire for user interface satisfaction are good enough. In addition, the two-factor interactions and the main effects: Type, Keyword, and Input significantly affect the degree of user satisfaction.

*Keywords*—User Interface Satisfaction; Mobile Map; Mobile Spatial Interaction; Touch Screen; Sense of Direction.

### I. INTRODUCTION

The functionality of mobile maps has been greatly increased by the use of new interactivity technology. Maps can be zoomed in and out and rotated without affecting the ratio of the display and can be easily combined with satellite images, aerial photographs and other sources of information to improve the user's understanding of the geographic database. The use of mobile maps is becoming popular with most mobile users; especially with the availability of free Google Maps for mobile phones. Google Maps for Mobile (GMM) claim that paper maps are obsolete. It offers street maps, a route planner for traveling by foot, car, or public transport and an urban business locator, for numerous countries around the world. Mobile phones can be used to search for local businesses, and then to obtain direction [3].

Touch screen tablet PC’s have become increasingly popular, since Apple launched the iPad device. They enable direct interaction with what is displayed on the screen, rather than indirectly with a mouse or touchpad. “*It looks like a giant iPhone,*” is the first thing users say when asked to test an iPad. However, from the perspective of interactive design, an iPad User Interface (UI) is more than a scaled-up iPhone UI [1]. The recent boom of popularity of mobile and tablet devices such as the iPhone and iPad open up a new world of

opportunity for mobile global positioning system (GIS) applications [13].

Literature concerning map-reading has shown that using a map is not an easy task for children, or even for adults [10]. Route-finding aids are important for finding routes in unfamiliar territory, in order to learn, about the surrounding environment. In particular, due to the advent of advanced information technologies, devices equipped with GPS receivers are becoming valuable tools for providing positional information [15]. With respect to navigational aids, studies such as in [7] have examined the effectiveness of GPS-based mobile navigation systems in comparison to paper maps and direct experience of routes. Their results show that GPS users travel longer distances and make more stops during the walk than map users and participants with direct experience of a route.

Various presentation formats for spatial information have been developed, including verbal navigational directions, static maps, interactive maps, 3-D visualizations, animations and virtual environments [12]. Dillemath [2] showed that a faster speed of travel and fewer navigation errors occur with a generalized map than with an aerial photograph. Some people can readily find their way back to a starting point along a route they have only experienced once, whilst others can only do this with considerable difficulty. There are large differences in individuals’ environmental spatial abilities [4, 8]. This individual difference between people is referred to as the sense of direction (SOD). Self-reported assessments of SOD have been found to provide quite accurate and objective measures of these abilities [5]. Participants with a good SOD (GSD) showed much better performance on route learning than those with a poor SOD (PSD). In addition, concerning gender differences in spatial cognition, it was suggested that males were superior to females in spatial information processing. Males preferred and used much better Euclidean spatial cues such as direction and distance, while females were likely to memorize landmark cues [9].

Usability is the extent to which a computer system enables users, in a given context of use, to achieve specified goals effectively and efficiently while promoting a feeling of satisfaction. Understanding what users expect to find and want to find, as well as what they typically use GMM for, can help the mapping service designers to provide a user-centered design. Moreover, understanding the needs of GMM will help improve the user experience and increase the service’s usability. The effects of four designing factors

including Size, Type, Direction Key, and Zoom function has been examined in [11]. The results indicate that participants with a better SOD would have the faster response time and would lower overall workload for target task. However, mobile maps differ from paper maps in that it provides a facilitation of spatial search. Keyword search using landmark or district and Chinese input methods using traditional Chinese hand-written input or traditional Chinese phonetic notation keyboard input for GMM spatial queries are concentrated on this research.

In this paper we discuss user interface satisfaction (UIS) that arose in using GMM on mobile devices with a touch screen interface such as the iPhone and iPad. This study differs from previous studies [7,10,11,12,15] in that it concentrates on the effects of keyword search and Chinese input methods that affect UIS. We first present our experimental design, including a description of the interfaces evaluated. Four design factors (Interface, Type, Keyword, Input) and two background factors (SOD and route knowledge) on UIS were examined. It helps Apps designers to provide an optimal user-centered interface for GMM. We follow with a description of our research methodology, define a classification scheme of SOD used in our analysis, and then present the results. The paper concludes with a discussion of design implications followed by future work.

## II. METHOD

### A. Participants

Forty-eight undergraduate, graduate students, teaching assistants and staff (24 females and 24 males) voluntarily participated in the experiment. Their ages ranged from 20.7 to 39.7 years old, with a mean of 23.7 years and standard deviation of 3.5 years. Nine out of Forty-eight participants had experience in using mobile E-maps other than GMM, before the experiment. They all had normal vision or corrected vision of at least 0.8 and no color-blindness. Participants were required to abstain from PC use for one hour before the formal experiment.

### B. Materials and Apparatus

The experiment used an iPad with a 9.7-inches multi-touch LCD display (1024×768 pixels) as a representative tablet PC and an iPhone with a 3.5-inch multi-touch retina display (960×640 pixels). An Optical Vision Tester was used to measure vision acuity and to test for color blindness. A digital video camera recorder (SONY DCR-PC330) was used to record the experiments and the post-experiment questionnaire. The luminance of experimental lab was 487~611 lux, as measured by a Lutron LX-101 Lux meter.

### C. Sense of Direction

Using the Santa Barbara Sense of Direction Scale (SBSOD) [5], 10 questions concerning spatial awareness and navigation allowed self-rating, using Likert's seven-point scale, before the formal experiment. Participants responded to each question by circling a number ranging from 1 (strongly disagree) to 7 (strongly agree). Four out of ten statements were positive, e.g., "My sense of direction is very

good," "I am very good at reading maps." The other six statements were negative, e.g., "I have a poor memory for where I left things," "I very easily get lost in a new city." These responses were reversed so that a higher score indicated a better SOD. The rating for SOD is calculated by summation of the scores for the ten SOD questions, as a SOD score and then these scores were categorized into two groups as SODG, using the median SOD as the divider for good SOD (GSD) and poor SOD (PSD). In addition, participants sat paper-and-pencil tests for route knowledge (RK), before the experiment. Their RK scores were recorded and categorized into two groups as RKG: good RK (GRK) and poor RK (PRK), based on the test result for route knowledge.

### D. Design of Experiment

This study seeks to provide an analytical model of usability of a GMM interface was evaluated using touch screen panels. The usability of the GMM interface was evaluated using a questionnaire for user interface satisfaction (QUIS), upon completion of three route-finding tasks. A four-way factorial design was used to assign each participant all three types of route-finding tasks to each participant and then a post-experiment questionnaire was used to determine user interface satisfaction. The four design factors consisted of: (1) Interface: tablet PC (iPad) vs. smart phone (iPhone), (2) Type: Electronic map (E-Map) vs. mixed map (M-Map, that is, E-map plus satellite), (3) Keyword: use landmark as keyword (Landmark) vs. use district (District), and (4) Input: traditional Chinese hand-written input (hand-written) vs. traditional Chinese phonetic notation keyboard input (phonetic notation). The illustration of the factors Type and Input is shown in Fig. 1. Demographical variables consisted of: (1) gender, (2) route knowledge (GRK/PRK), and (3) sense of direction (GSD/PSD). Forty-eight participants (24 female and 24 male) participated in this experiment. Three route-finding tasks were assigned to each participant; (1) find targets, (2) identify cardinal directions and (3) identify the approximate distances. Cardinal directions were based on 8-sectors model (North, East, South, West, North-East, South-East, South-West and North-West), while approximate distances corresponded to a set of ordered intervals, where the order of symbolic distance values describes distances from the nearest to the furthest [6]. The time to correctly complete the target task, the time to correctly complete the direction task and the time to correctly complete the distance task using a GMM interface were measured (omitted due to limitation of paper length).

The study areas were northern, central and southern metropolitan district of Taiwan, in Taipei, Taichung, and Kaohsiung Cities, respectively. The participants all started from the same point. The mapped area was dynamically updated as the user moved in space.

### E. Questionnaire for User Interface Satisfaction (QUIS)

The questionnaire for user interface satisfaction (QUIS) is a structured assessment of usability. It is useful in the early stages of the development of a user-centered design. The International Organization for Standardization (ISO) defines

the usability of a product as “the extent to which the product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.” The usability of interactive products is generally defined by the ease with which they can be learned, their effectiveness in use, and the extent to which the user finds them enjoyable to use [14]. In this study, usability is defined by the following parameters: pleasure, interactivity, efficiency, ease of use, ease of recovery from error, memorability, satisfaction, and recommendation.

#### F. Procedure

At the beginning of the experiment, naive participants were asked to familiarize themselves with the GMM interface. They practiced using the device until they were proficient. Prior to the formal experiment, they were asked to complete an individual background questionnaire, including self-rating of sense-of-direction and route knowledge. One of sixteen combinations of treatment was randomly assigned to one of the participants and the participants were subjected to the route-finding tasks. After the completion of all three route-finding tasks, each participant completed the QUIS using a 10-point Likert’s scale.

### III. RESULTS AND DISCUSSION

#### A. Descriptive Statistics and correlation matrix

The overall user interface satisfaction (UIS) score is calculated by summation of the scores for the eight UIS questions as the UIS score. The descriptive statistics are shown in Table 1. The mean UIS for females (54.0) is higher than that for males (50.6). The participants with GRK have a higher mean UIS (59.4) than those with PRK (46.8). The mean UIS for GSD participants (54.1) is higher than for those with PSD (50.5). The mean UIS for using iPad (56.8) is higher than that for using iPhone (47.8). The mean UIS for using E-Map (55.1) is higher than that for using M-Map (49.5). The mean UIS for using landmark search directly (54.0) is higher than that for not using landmark search, but using district search (50.7). The mean UIS for using hand-written keyword input (55.4) is higher than that for using phonetic notation input (49.2). Based on the eight ordinal-scale items of QUIS, the results of Spearman’s rank correlation between pairwise items are shown in Table 2. It can be seen that there is a statistically significant and a moderately positive correlation between pairwise items.

#### B. Reliability and Validity

Cronbach’s alpha is a coefficient of reliability (or consistency). It is a measure of internal consistency, that is, how closely related a set of items are as a group. A “high” value for alpha is often used (along with substantive arguments and possibly other statistical measures) as evidence that the items measure an underlying (or latent) construct. The results of QUIS in this study indicate a value of Cronbach’s alpha is 0.925 and there is good internal consistency. However, a high alpha does not imply that the measure is one-dimensional. If, in addition to measuring internal consistency, evidence is required that the scale in

question is one-dimensional, additional analyses can be performed. Exploratory factor analysis (EFA) is one method of checking dimensionality. The results of EFA for the QUIS indicate that 66.8% of the total variation is explained by only one common factor, which is named as the degree of user interface satisfaction (DUIS). Examination of the goodness-of-fit indices confirms that the confirmative factor analysis (CFA) model has been well designed.

#### C. Individual differences in UIS

The relationship between SODG, RKG, Gender and UIS were investigated for the three tasks. The two-factor interaction plot is shown in Figure 2. It indicates that females with PSD tended toward a higher UIS than males with PSD; however, no significant difference is evident between females and males in the GSD groups (Fig. 2(a)). Similarly, females with PRK have a higher UIS than males with PRK; both females and males with GRK have nearly the same UIS, on average (Fig. 2(b)). In addition, GRK groups with PSD and GSD have a higher UIS than PRK groups (Fig. 2(c)). There is insufficient evidence of statistical significance (all  $p > 0.05$ ) for the two-factor interactions.

#### D. Analysis of variance and Interaction plots

Based on the results of ANOVA in Table 3 indicate that the significance of the two-factor interaction of Interface\*SODG and Interface\*RKG ( $F=7.151$ ,  $p=0.011$  and  $F=4.323$ ,  $p=0.045$ , respectively) are all supported. The main effects: Type, Keyword, Input, and RKG are statistically significant ( $p < 0.05$ ). To interpret the interaction, comparisons between GSD and PSD groups depend upon whether they use iPad or iPhone. In Fig. 3(a), the PSD group using iPad tended toward a higher UIS than the GSD group, but the PSD group using iPhone tended toward a lower UIS than the GSD group, on average, for GMM. PSD participants prefer using iPad to iPhone, but for iPhone users, GSD participants have a higher UIS than those with PSD. It is interesting to note how sense of direction affects the satisfaction users of the different interfaces. In Fig. 3(b), the GRK group using either iPad or iPhone tended toward a higher UIS than the PRK group, but the PRK group using iPad for GMM tended toward a higher UIS on average, than those using iPhone. Similarly, PRK participants prefer using iPad to iPhone. However, for iPhone users, GRK participants have a higher UIS than the PRK group.

#### E. Discussion

Kato and Takeuchi [9] argue that individual differences in wayfinding strategies between GSD and PSD female undergraduate participants. GSD participants showed much better performance on route learning than PSD participants [9]. Similarly, our results show that GSD participants have higher UIS than PSD participants on using GMM. While little work has explicitly considered how GRK/PRK groups interact with GMM on iPhone/iPad for UIS. A significant amount of research has explored how route knowledge is best conveyed by both interfaces used in our study. Moreover, a significant effect also has explored how SOD is best communicated with both interfaces. Effective route

maps must provide information that is necessary and sufficient to make the right choice at each decision point [16]. Agrawala and Stolte [17] argue that for maps on mobile devices it is particularly important that the routes are simplified and extra information is removed. It is consistent with our study that E-map has superior UIS than M-map. Reliability of the questionnaire was tested on 532 undergraduates as participants using internal consistency and split-half methods in Kato and Takeuchi's study [9]. A moderately high reliability was obtained. Similarly, both the reliability (Cronbach's alpha) and confirmative factor analysis model of the questionnaire for user interface satisfaction are good enough in this study. However, only forty-eight participants used in the experiment.

IV. CONCLUSION AND FUTURE WORK

The results of this study have implications for mobile spatial interaction in general. Most GMM users prefer using E-map to M-Map. They also prefer using landmark as keyword to district and prefer using traditional Chinese hand-written input method to traditional phonetic notation input. The poor SOD (PSD) group prefers using iPad to iPhone. The poor RK (PRK) group prefers using iPad to iPhone and the good RK (GRK) group has a higher UIS than the PRK group. It is also important to integrate the impact of the design factors and individual differences on the user performance of mobile spatial interaction. The results of quantitative measurements and subjective assessments will be used as the guidelines to provide a better solution and to meet the demands of usability for mobile spatial interaction applications.

ACKNOWLEDGMENT

This work was supported by the National Science Council, Taiwan under Grant NSC 99-2221-E-029 -029.

REFERENCES

[1] R. Budiu and J. Nielsen, Usability of iPad Apps and Websites, 2nd ed., <http://www.nngroup.com/reports> <retrieved: Jan. 2012>.  
 [2] J. Dilleuth, "Map Design Evaluation for Mobile Display," Cartography and Geographic Information Science, vol. 32, pp. 285-301, 2005.  
 [3] Google Maps, <http://m.google.com.tw/maps> <retrieved: Jan. 2012>.  
 [4] M. Hegarty, D. R. Montello, and A. E. Richardson, T. Ishikawa, and K. Lovelace, "Spatial Abilities at Different Scales: Individual

Differences in Aptitude-Test Performance and Spatial-Layout Learning," Intelligence, vol. 34, pp. 151-176, 2006.  
 [5] M. Hegarty, A. E. Richardson, D. R. Montello, K. Lovelace, and I. Subbiah, "Development of a Self-Report Measure of Environmental Spatial Ability," Intelligence, vol. 30, pp. 425-447, 2002.  
 [6] J. H. Hong, M. J. Egenhnfer, and A. U. Frank, "On the Robustness of Qualitative Distance-and Direction Reasoning," In: Peuquet, D. (ed) Proceedings of Autocarto, vol. 12, 1995.  
 [7] T. Ishikawa and H. Fujiwara, O. Imai, and A. Okabe, "Wayfinding with GPS-Based Mobile Navigation System: A Comparison with Maps and Direct Experience., Journal of Environmental Psychology, vol. 28, pp. 74-82, 2008.  
 [8] T. Ishikawa and D.R. Montello, "Spatial Knowledge Acquisition from Direct Experience in the Environment: Individual Differences in the Development of Metric Knowledge and the Integration of Separately Learned Places," Cognitive Psychology, vol. 52, pp. 93-129, 2006.  
 [9] Y. Kato and Y. Takeuchi, "Individual Differences in Way Finding Strategies," Journal of Environmental Psychology, vol. 23, pp. 171-188, 2003.  
 [10] L. S. Liben, K. A. Kastens, and L. M. Stevenson, "Real-World Knowledge Through Real-World Maps: A Developmental Guide for Navigating the Educational Terrain," Developmental Review, vol. 22, 267-322, 2002.  
 [11] Y. L. Lin and C. H. Wang, "Evaluation of Wayfinding Performance and Workload on Electronic Map Interface," J.A. Jacko (Ed.): Lecture Notes in Computer Science, Human-Computer Interaction, Part III, HCII 2011, vol. 6763/2011, pp. 311-320, Springer-Verlag Berlin Heidelberg, 2011.  
 [12] D. R. Montello, D. Waller, M. Hegarty, and A.E. Richardson, "Spatial Memory of Real Environments, Virtual Environments, and Maps," In: Allen, G.L. (ed.) Human Spatial Memory: Remembering Where. Mahwah, NJ: Erlbaum, pp. 251-285, 2004.  
 [13] H. Samet, B. E. Teitler, M. D. Adelfio, and M. D. Lieberman, "Adapting a Map Query Interface for a Gesturing Touch Screen Interface," Proceedings of the 2011 International Conference on the World Wide Web, vol. 2, pp. 257-260, 2011.  
 [14] S. Sharp, Y. Rogers, and J. Preece, Interaction Design: Beyond Human-Computer Interaction, 2nd ed., John Wiley & Sons, Ltd., 2007.  
 [15] N. Shoval and M. Isaacson, "Application of Tracking Technologies to the Study of Pedestrian Spatial Behavior," Professional Geographer, vol. 58, pp. 172-183, 2006.  
 [16] M. P. Daniel and M. Denis, "The Production of Route Directions: Investigating Conditions that Favour Conciseness in Spatial Discourse," Applied Cognitive Psychology, vol. 18, pp. 57-75, 2004.  
 [17] M. Agrawala and C. Stolte, "Rendering Effective Route Maps: Improving Usability Through Generalization. In: Proceedings of SIGGAPH, Los Angeles, California, August, 241-250, 2001.



Figure 1. iPhone screen shots for GMM: (a) using E-Map, (b) using M-Map, (c) keyword search using traditional Chinese hand-written input method, and (d) keyword search using traditional phonetic notation input method (<http://m.google.com.tw/maps>).

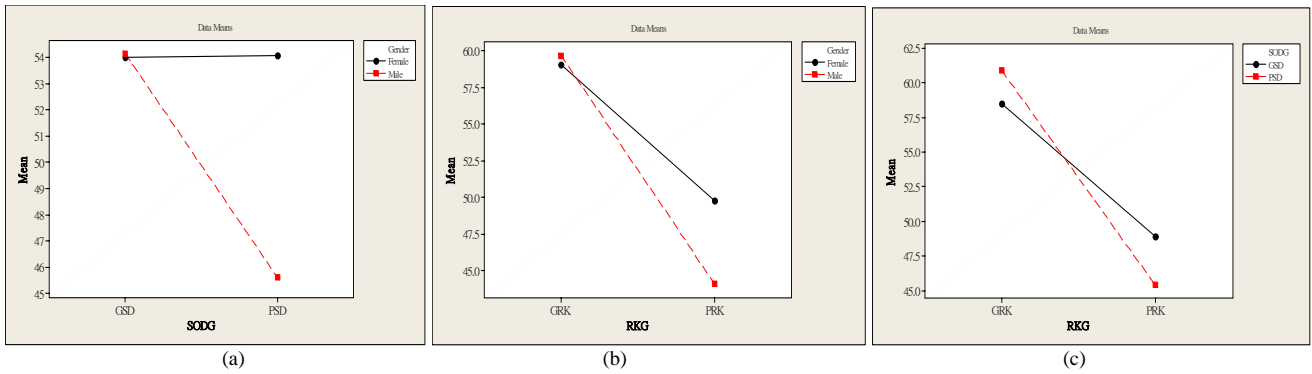


Figure 2. Interaction effects plot for UIS: (a) Gender\*SODG, (b) Gender\*RKG, (c) SODG\*RKG.

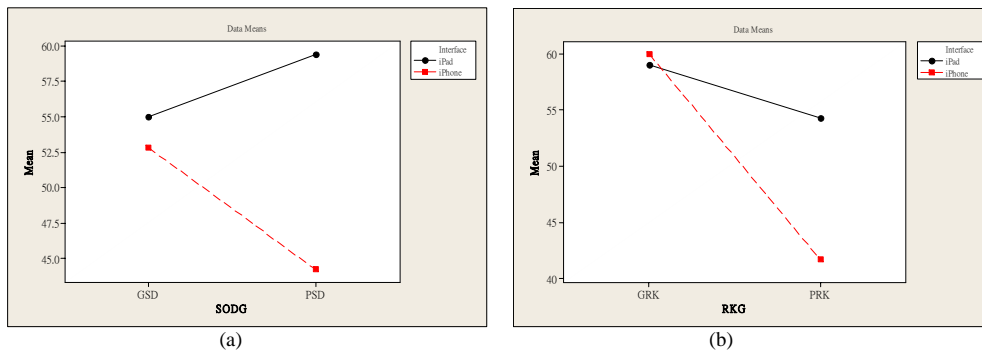


Figure 3. Interaction plots for UIS: (a) Interface\*SODG and (b) Interface\*RKG.

TABLE 1. Descriptive statistics of user interface satisfaction (UIS) for the GMM

Variable	Level	n	Mean	StDev	Min	Q1	Median	Q3	Max
Gender	Female	24	54.0	12.85	20	45.50	53.00	66.00	74
	Male	24	50.6	16.71	19	35.00	58.00	63.00	68
Sense of Direction	GSD	24	54.1	15.10	19	48.00	61.00	64.50	74
	PSD	24	50.5	14.71	21	39.25	52.50	62.25	70
Route Knowledge	GRK	21	59.4	10.97	27	54.00	62.00	67.00	74
	PRK	27	46.8	15.31	19	35.00	49.50	61.00	70
Interface	iPad	24	56.8	11.5	27	51.00	60.00	66.75	70
	iPhone	24	47.8	16.61	19	35.50	51.50	61.00	74
Type	E-Map	24	55.1	12.35	27	47.00	56.50	66.00	74
	M-Map	24	49.5	16.79	19	35.50	55.50	62.50	69
Keyword	Landmark	24	54.0	13.30	20	47.50	57.50	63.00	70
	District	24	50.7	16.38	19	39.25	54.00	66.00	74
Input	Hand-written	24	55.4	12.27	24	47.50	57.50	65.25	74
	Phonetic Notation	24	49.2	16.74	19	36.00	55.50	63.00	70

TABLE 2. Spearman’s rank coefficient of correlation of UIS for GMM

Coefficient of correlation (P-Value)	Pleasure	Interactivity	Efficiency	First priority use	Ease of use	Error recovery	Memorability	Satisfaction
Pleasure	1	0.725	0.574	0.644	0.71	0.549	0.427	0.607
		.000**	.000**	.000**	.000**	.000**	.002*	.000**
Interactivity	0.725	1	0.752	0.597	0.63	0.61	0.479	0.64
	.000*		.000**	.000**	.000**	.000**	.001*	.000**
Efficiency	0.574	0.752	1	0.683	0.548	0.614	0.454	0.666
	.000**	.000**		.000**	.000**	.000**	.001*	.000**
First priority use	0.644	0.597	0.683	1	0.647	0.466	0.538	0.746
	.000**	.000**	.000**		.000**	.001**	.000**	.000**
Ease of use	0.71	0.63	0.548	0.647	1	0.668	0.389	0.66
	.000**	.000**	.000**	.000**		.000**	.006**	.000**
Error recovery	0.549	0.61	0.614	0.466	0.668	1	0.466	0.561
	.000**	.000**	.000**	.001**	.000**		.001**	.000**
Memorability	0.427	0.479	0.454	0.538	0.389	0.466	1	0.618
	.002**	.001**	.001**	.000**	.006**	.001**		.000**
Satisfaction	0.607	0.64	0.666	0.746	0.66	0.561	0.618	1
	.000**	.000**	.000**	.000**	.000**	.000**	.000**	

\* P-value <0.01

TABLE 3. ANOVA of UIS for GMM

Source of Variation	Sum of Square	DF	Mean Square	F	P-value
Interface	277.42	1	277.42	2.057	0.16
Type	671.621	1	671.621	4.98	.032*
Landmark	590.062	1	590.062	4.375	.043*
Input	1110.106	1	1110.106	8.231	.007**
Gender	226.136	1	226.136	1.677	0.203
SODG	273.547	1	273.547	2.028	0.163
RKG	2016.031	1	2016.031	14.949	.000**
Interface*Gender	73.199	1	73.199	0.543	0.466
Interface*SODG	964.403	1	964.403	7.151	.011*
Interface*RKG	583.021	1	583.021	4.323	.045*
Error	4989.927	37	117.714		
Corrected Total	10368.3	47			

\*p<0.05; \*\*p<0.01

## Tablet PCs – An Assistive Technology for Students with Reading Difficulties?

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**—This paper presents the results of a descriptive case study concerning adoption of iPad or other tablets as assistive technology. Two pilot studies concerning the adoption and use of the iPad for active reading in a teaching/learning situation have recently been conducted at elementary school level and at university level. In the course of both studies, students with reading difficulties were encountered. For each group of students, a key case has been chosen. The paper presents our findings regarding adjustments that needed to be made for these students and initial research on iPad usability for students with special education needs. By describing two instances, one involving a university student and the other two elementary school children, we hope to bring attention to application of ICT for students with reading difficulties. Students with this kind of impairment are often neglected in comparison to students with visual impairments or other disabilities. In one case, the iPad has been successfully integrated into students' life as an assistive technology. The cases may be both instructive and inspirational for educational situations involving students with similar disabilities as adjustments and applications used to help students do not involve any large investments in software or devices.

**Keywords**-assistive technology; iPad; reading difficulties; tablet PC; technology adoption

### I. INTRODUCTION

Two pilot studies involving the use of iPads for active reading in a teaching/learning situation have recently been conducted [1], [2] and [3]. One of the studies has involved University students and the other, 4<sup>th</sup> grade children in elementary school. The goal of the studies was to explore the potential of mobile and wireless technologies (the iPad was used in both cases) to change classroom information ecologies and enable anytime anyplace learning situations. The concept of information ecology was introduced by Nardi and O'Day in [4] as “a system of people, practices, values, and technologies in a particular local environment” and it focuses on five defining characteristics 1) it is a system 2) the system contains diversity of people and tools 3) there is a change or co-evolution happening over time and through use of technology 4) keystone species are part of the ecology (their presence is critical for the system's survival) and 5) local habitation (the habitation of technology is its location within a network of relationships). Nardi argues: “Human expertise, judgment and creativity can be supported, but not

replaced by computer-based tools.” One should contemplate the technology with both the head and the heart and not fall prey to either technophilia or technophobia. The approach based on classroom information ecology was taken in all the studies we conducted with iPads.

In [5], Turkle states the following regarding computer technology: “Most considerations of the computer concentrate on the “instrumental computer”, on what work the computer will do. But 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, especially the way we think about ourselves. We saw both the view of “instrumental tablet” and “subjective tablet”.” The instrumental side answering questions around how the iPad may be best used in the classroom and outside of the classroom for the purposes of learning. The subjective side addressing the plethora of factors such as personal relationship with iPad, social changes it induces, taking a larger freedom in designing the curriculum (empowerment), avoiding stigmatism in cases of children with special needs, self-image, changes in a way of thinking and interacting with technology etc.

The personal relationship to the device opens up for the new uses of the iPad in the classroom setting, such as assistive technology (AT) for children with special needs (see, for example, [6]).

The adoption of tablet PC as an assistive technology became the subject of our case study that was conducted from September 2010 until September 2011.

In this paper, we will focus on two cases that we have worked with, involving students with reading difficulties. The first one is a University student whom we will call Mary in this report. Mary is diagnosed with dyslexia. She was highly motivated on her own to work with this new possibility iPad offered.

Our second case presents a much more sensitive situation, involving two elementary school children, whom we will call Iris and Josh in describing their case. Iris and a Josh are both aged nine. These children do not have any kind of diagnosis. Neither the teacher nor parents are trying to have the children diagnosed. This approach may have some advantages such as protecting the children from stigmatization, but it also has problems. The main issue is children's low self-esteem [7, 8]. Iris, in particular, has a twin sister at school, with no impairments. On the contrary,



the sister excels in academic achievements. Iris is aware that she does not perform as well as her sister. Josh may or may not be aware that he has a problem at all. So this case is extremely interesting for bringing to light issues related to awareness of the existence of the problem and its acknowledgment that can have effects on the policy making for schools regarding AT, rather than silencing it until the students are older.

While the causes of dyslexia may be complex to understand, it often manifests as a learning disorder marked by impairment of the ability to recognize and comprehend written words. A significant proportion of students seeking help from the Accessibility Services at the University of Oslo are diagnosed with dyslexia. About 46% of those [9,10] are first diagnosed after they start their higher education. This fact is in stark contrast to other kinds of disabilities such as visual impairment, hearing loss or physical impairments, which are usually diagnosed much earlier. This is important both for how fast and how much help these students get in order to minimize the impact of their impairment on learning [11]. A further complication with offering help to dyslexic students is that they often require individual adjustments.

In parallel with technological developments, from hardware to Internet and Communication Technologies (ICT), systems for dyslectic and visually impaired users have been developed. The effects of these systems have been reported in a number of different studies [12]. For instance, it has been found that different voices reading the text have an impact on the user's text comprehension [13].

School and university libraries have traditionally offered help for their users with dyslexia or visual impairments. Libraries would often have expensive equipment such as a special enlargement screen and computers using "text to speech" software. The reasons the libraries have all this special equipment is the prohibitive cost of the equipment paired with "access for all" philosophy. An additional benefit for users was help with mastering of this rather complex equipment generously provided by library personnel.

With the arrival of the e-book readers, and later tablet PCs, this scene is changing for dyslexic students. What they had to go to the library for before, they could now have with them, anywhere, anytime. A new world of possibilities has opened up for the dyslectic community, although tablets may be used as assistive technology (AT) for other kinds of impairments as well [14].

In the first case, Mary cooperated with us in trying different approaches, meandering between problems and solutions until we found what works for her. After more than a year of following Mary's iPad use, we can report that this technology has made a significant difference in her academic performance and she became somewhat of a virtuoso in handling her iPad. In [15] the author states: *By using the very digital media that is helping drive this information society, computing technologies may be a viable means of providing reading support and accommodation. For such technologies to be successful, though, they must be adopted into regular use. Unfortunately, studies have shown that 35–50% of all assistive devices are abandoned after purchase.* Mary has made a margin of those who keep on using their AT.

In the second case, such an open approach was not possible. Instead, an experiment involving the two children with reading difficulties and a control group was carried out. The experiment helped us showcase the potential the iPad (or another tablet) may have as an AT for children with reading difficulties.

Both cases suggest strongly that at least some portion of dyslexic students could be helped by similar means. In [15] it is pointed out that many factors (socio-cultural, technical, economic, environmental etc) influence adult adoption of assistive technology. We find that all those same factors influence the children. Invisible nature of reading disorder and even stronger impulses not to disclose it are of huge importance with children. The statistics valid for adult population (5-15%) are probably the same within the young population, except that these are not available, in part due to the invisibility issue. Therefore it is of large importance to bring awareness to this situation. Some of the methods developed for adults such as Value Sensitive Design proposed in [14] may be of use with children as well. The mobile AT adds additional value in that it does not draw attention to the person using it (as opposed to sitting in a special room in the library, in front of a huge screen). Someone "reading" the text by listening to it from the iPad looks quite "normal", but the impact of this kind of AT may be quite huge on children's education, social life and self-esteem among other things.

The technology adoption issues are difficult in the best of circumstances with so many factors influencing the success or failure. Assistive technologies are even more difficult. However, from 2020, universal accessibility will be enforced by law in Norway, and thus it is timely to investigate how this can be done in the classroom.

From the design perspective, solutions found for groups with special needs often find their way to the mainstream. In this case, tablets may become an example of a device designed for the mainstream, but having potential to be accepted and adopted by groups with special needs.

## II. THE APPROACH TO THE PROBLEM

The research conducted around the two pilot studies [1,2,3] has organically led to discovery of students with disabilities. We, at the beginning of the pilot studies, did not have any intention to study the use of iPad by students with dyslexia, but we thought that seeing what Mary does with the tablet will be very interesting. We were going to interview her periodically and record what happens. However after the very first interview with Mary, we realized that we would need to take an active role in making adjustments for her, as well as observing her in the class, having interviews both with other stakeholders, such as software producers or anyone else who could help her. We needed to think of what kinds of software, applications as well as potentially other devices would work for her. We also quickly found out that it is fascinating to learn about learning practices of dyslexic students. And so it also unfolded into looking at policies including privacy, role of environment, social and cultural positioning of the student, and the role of teachers. This paper will cover only the

grounds of how the technology became part of Mary's everyday life, and how working with her inspired us to look at the reading difficulties in younger children. We find the problem of adoption of technology in the case of young children even more challenging, due to the fact that they themselves are often not aware that they have difficulties with reading and learning that can be helped.

The nature of the problem that we were looking at was such that only a few subjects were available for the study. Thus, naturally, a case study (see for example [16]) became a method of choice. We use a descriptive case study, an in-depth study of a specific instance of assistive technology adoption, with explanatory purpose, placing the spotlight on what might become important to look at more extensively in the future research. Our techniques have included direct observations in class, interviews with Mary, and interviews with elementary school teacher as well as Iris's family.

### III. MARY'S CASE

Mary was appointed by the University of Oslo Accessibility Services to participate in a larger project, involving introduction of the iPad into a geology course [3]. In practice, this meant that Mary received an iPad to use. She did not get any special support with it, except for having Dropbox, iAnnotate, and a 25\$ gift card as part of the iPad deal.

Some major problems that she encountered while attempting to use the iPad to her advantage, were disclosed during our first interview with her. This was the story in a nutshell: since Mary is dyslexic, the Norwegian Library for the Impaired (NLB) was charged with the task of finding, if available on the market, curriculum for her in speech synthesis. Usually, the curriculum is delivered in the Daisy file format, which contains both speech and text. Using special software, Mary should be able to hear the text and watch it being highlighted at the same rate as the speech is progressing. However, the Daisy file reader was made for the Windows platform and Mary is a Mac user. The NLB could not locate the software for Mac or iPad. On her own initiative and without help, Mary tested several free applications from the Appstore that can read Daisy files on the iPad. But she ran into problems again. The Apps would crash all the time. She thought the problem was caused by lower quality of the free software, and thus, she bought a \$30 full price version of the Daisy reader. There was no improvement. The program kept crashing. With some evident frustration, she shared: *"So I tried some different software that worked a little, but it froze often, both audio and text, and sometimes the iPad went completely dead!"*

There were other technical problems contributing to this negative overall experience. An example is that the student housing where she lives, has no Wi-Fi connection. *"If I had an Internet connection I would have used it (the iPad) more actively"*. In summary, Mary was not really able to use the iPad in the ways she wished and needed to do.

We made a joint agreement to give it another try during the following semester (Spring 2011). One of the authors started investigating the problems with the Daisy files. The application Mary bought to play Daisy files with on the iPad

was Voice of Daisy (VOD). After testing it on another iPad with no improvement, a request for information on the files and plea for help was sent by email to NLB. Although the library showed huge interest in our approach, the only information we got was how the CDs with Daisy files were produced in the house. A second effort was then made involving contact with the Japanese developer of the VOD application. After several emails, they resolved the problem. The reason for crashing of the App was the poor quality and the erroneous offset of the files she received from the NLB.

As part of the agreement between the student and us a supplementary intensive support period was given to her, teaching her how to use the iAnnotate and other iPad applications. Mary also agreed on monthly interviews with the authors, in order to make sure that the progression of the use of the iPad was not interrupted by yet another technical problem.

It is very interesting to note that we have asked if we could observe her working with the iPad in her courses. Mary at once agreed to be observed at the lectures, with many students attending. However, she definitely did not want to have anyone observing her at small work group meetings that are part of the course set up: *"If I let you do that, then I for sure will not make any friends in this class."* Mary spoke directly about the issues of stigmatization in relation to her need for assistive technology.

During the next few meetings in the spring semester she reported increased use of the iPad for studying. And then, one day it all fell into place. With some help, she had developed her own way of working with the iPad, turning it into a proper AT tool. She was able to use it anytime, anywhere. And most importantly, she really enjoyed it!

Mary was using different applications for different needs. Voice of Daisy was used for the part of the curriculum involving books. For articles she used two applications simultaneously. The first one was Speak-It with the possibility to cut and paste part of the text and hear it. The second one was iAnnotate where she could mark the text, annotate it and enlarge it while simultaneously listening to it. She was very pleased with the fact that she could choose part of the text she wanted to hear, and was not forced to listen to the whole text. She found her "own" special ways to use the color and the strikethrough (see Figure 1).

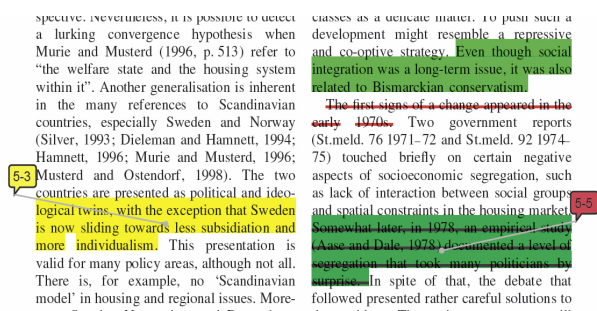


Figure 1. Example of how Mary worked with text while listening to it.

The color was used to group similar topics, while the strikethrough was used to remove uninteresting parts of the text. The tactile interface made quick selection of the text

possible. Using a normal laptop and mouse interaction would have taken much longer. Mary's interaction with the iPad also became beautiful to watch; her movements are quick, certain, effective, lightly dancing around the touch surface (see Figure 2).

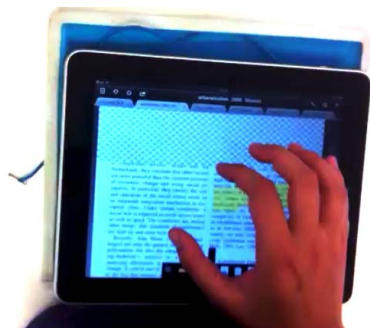


Figure 2. Mary is engaging with the text before it even loads fully.

Using VOD on the iPad was also easier than on the PC. She explains: "Zooming in and out gives me a better view. For example, I can always quickly find out where I am in the text."

In the classroom, she had the possibility of taking and grouping all her notes on the iPad, using the default software such as Notes shown in Figure 3.

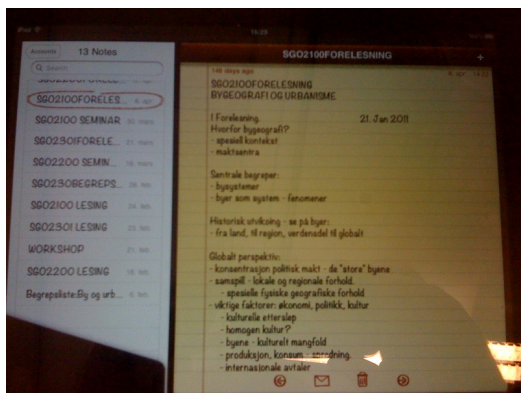


Figure 3. Taking notes in class on the iPad

The aftermath of these joint efforts to make the iPad into an AT shows that Mary's interest in her field has increased; her confidence in being able to finish her studies has increased and her overall attitude towards AT has improved.

#### IV. IRIS'S AND JOSH'S CASE

In the beginning, the children with reading difficulties were indistinguishable from their peers. Together, they made up a class of 26 students in a rural Norwegian school. Six iPads were given to the class, five for students and one for the teacher. Classroom got wireless connectivity in conjunction with the pilot study, thus enabling wider use of the Internet. Digitalized curriculum is not yet common in elementary schools. In spite of this, access to digitalized curriculum was obtained from the academic publisher (free of charge) for Religion Studies, Mathematics and Science.

English is relevant both as the subject at school and as the language of applications. The students have some knowledge of the language, but many are far from fluent. The traditional way of teaching English was supplemented from the start of the study with stories and Apps (such as Alice in Wonderland or balloons) that could help students to improve their English through play. One day per week was set as an observation day.

All of the children were rather excited about having the iPads in the classroom. They could also bring them home (according to the schedule they made).

From the very start, we observed that many pupils liked enlarging text, sometimes quite a bit (see Figure 4), while reading.

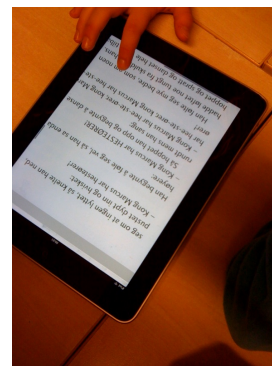


Figure 4. Iris is enlarging text while reading

It is worthwhile mentioning a study [17], done on Kindle, but with similar possibilities to enlarge letters. The authors report: "The Kindle provides a choice of six different font sizes. During this study, Amy generally kept her font at a larger size than Winnie. In an interview, she explained that it helped her "read faster when the text was large." The varying text size did create some challenges on days when the girls decided to partner read, as the visual layout of their Kindle "pages" differed. The girls quickly learned to synchronize their settings when reading together."

The preference for larger text and ease in reading with large letters is related to dyslexia [18]. If a child enlarges the text to an unusual size, it signals that the child may have reading difficulties. Both Iris and Josh do prefer to read on the iPad to reading from paper, mainly because of the ability to enlarge the text.

When iPads were collected for the first time in order to see what kind of content the children have placed on their them, one iPad differed from others significantly (see Figure 5). Josh has organized all the content into thematic groups, being displayed quite neatly on the iPad. That was strange, but even stranger was the fact that one of those groups had to do with languages and translating from one language to another using speech.

The organization of content soon became a class standard, but no other students ever installed apps for learning languages or translating from one language to another. These actions made Josh visible to us.



Figure 5. Josh’s iPad, compared to another one from the class. Note: the third item in the first column is a category marked as snakk (trans. speak), containing language speech based applications

In the course of the pilot study [1], Iris and the rest of her family were interviewed twice. During the first interview, the parents pointed out that the girls are “much more” interested in homework, especially the one of them that has some problems with reading. This is how we got interested in Iris.

In a later interview with the teacher, it was confirmed that both children have difficulties reading. However, the teacher told us: “*This is very, very confidential.*”

The teacher had some specific wishes regarding the use of the iPad in class as assistive technology for children with special needs. She said: “*Groups who need special education can be helped by the school. Having some iPads could be like “a carrot on the stick” for students who cannot be helped so easily and who are struggling a lot.*” (trans. Culén). The statement, positive as it is, also had a note of resignation in the face of the complexity of the problems comprising the child’s self-esteem, self-perception, perception by others (often involving stigmatization), parents’ involvement etc. On top of these challenges the school would also face organizational challenges around supporting the adoption for everyday use coupled with adaption to the needs of an individual student.

The teachers comment about confidentiality has made direct inquiry with children impossible, as well as use of the same approach as we had with Mary. We were limited to direct observations and questions around why do students like to use the iPad.



Figure 6. The children participating in an experiment comparing the understanding and retention after reading from paper and ipad using SpeakText program.

In order to be able to grasp what kind of difference in comprehension the iPad (through text-to-speech App

SpeakText, see Figure 6) could enable for these two children we designed a simple experiment. The experiment engaged five children: the two with reading difficulties, and 3 without difficulties, including Iris’s twin sister. The purpose of the experiment was to give an indication of what could be done with the use of iPad, and not to provide any statistically significant results. The real purpose of the experiment was not presented to the students for confidentiality reasons. Rather, the experiment was presented as testing of the effectiveness of reading on the iPad.

A. Experiment design

Our null hypothesis was that there is no difference in understanding the text for children with and without reading difficulties when they read from paper and when they select the text on iPad and heard it read to them. It involves two independent variables each having two conditions (children with and without reading difficulties and reading from paper or iPad’s app SpeakText).

The dependent variable (understanding of the text) was measured by how the children answered 8 simple questions after the reading (or hearing the text). Four of the questions were retention (memory) based and the other four based on understanding causes and effects in the story.

Due to the small number of children with reading difficulties that we could recruit, the within the group design was an obvious choice. Thus each student repeated the reading session, followed by the answering session, twice – once with paper and once with the iPad, where which was to be done first was determined at random (see Figure 6).

The reading was done from two distinct passages from the same text, approximately equal in length (374 words vs. 380 words), from a children’s book. The iPad app SpeakText, with voice over text and highlighting while the text is being read, needed 3 min. and 18 seconds to read the 380 word paragraph. The children’s reading from the paper based on 374 words was timed for all five children. The clock was stopped when the child indicated that they have finished reading the paragraph. In one case, there is a slight imprecision due to the fact that we did not stop the watch precisely enough when the child indicated that she finished the reading. The result is given as an approximate time in Table 1.

The answer session was not timed, but the children knew what to expect the second time around and they were somewhat faster on the second set of questions than on the first, indicating that some learning effect has taken place.

B. Results

Table 1 summarizes the results obtained from children without reading difficulties. Table 2 summarizes the results from those with reading difficulties. Each field in the table gives the number of questions that were answered correctly by the child. As mentioned above, in conjunction with paper reading, the reading time was recorded (the iPad time was always 3 min. and 18 seconds). After the reading was over, the questions were handed in and the students could no longer view the text they just read (or heard).

TABLE I. SUMMARY FOR CHILDREN WITHOUT READING DIFFICULTIS

Under- standing	Child 1	Child 2	Child 3
<b>iPad</b>			
Memory	4	4	4
Comprehension	4	2	3
<b>Paper</b>	Time 1:54	Time 2:27	Approx. 2:30
Memory	3	3	4
Comprehension	4	2	3

TABLE II. SUMMARY FOR CHILDREN WITH READING DIFFICULTIES

Understanding	Child 1	Child 2
<b>iPad</b>		
Memory	3	4
Comprehension	2	4
<b>Paper</b>	Time 15:29	Time 6:27
Memory	2	1
Comprehension	0	0

Note that neither of the children with reading difficulties has answered any comprehension questions when reading from the paper. Retention questions did not fare much better. Although not perfect for both children, the results after the iPad use were improved. It is also interesting to note that for the children without reading difficulties, the iPad use shows slightly better results.

These results are, of course only indicative due to a very small sample size.

Post experiment, we collected impressions from the children around the experience of reading from the iPad. The children remarked that they liked zooming on the text as well. Iris in particular mentioned twice that zooming helps her. It would have been perhaps interesting to repeat the experiment with both readings from the iPad, one of them with possibility of enlarging the text and the other one with SpeakText App.

### V. CONCLUSION

In the process of working with the two cases, we believe to have seen how tablet PC can bring forward some new possibilities as AT in this sensitive and complex field.

In spite of the small sample size, the case of elementary school children, at the very least, indicates the need for more research related to AT. Our hope is that larger studies will be conducted at elementary schools worldwide, inspired by this, and similar small studies. We view the outcome of this study to also be a contribution to the body of evidence that mobile technology may be in some cases effectively used as AT.

The introduction of the iPad in the elementary class has, in general, been a success. In particular, it offers clear support to some children with reading difficulties. Even in the situation where the impairment itself is confidential and kept in silence. While it is still true that each child/student with dyslexia or with reading difficulties needs individual assessment as to what works and what does not work, there is a number of possibilities that were very simple to try with iPads (things like different apps for text to speech, changing

synthetic voices in order to find the one that works the best, enlargements, color annotations etc).

A very important point in favor of mobile assistive technologies is that it minimizes stigmatization for the ones using it. For example, for Mary, she could sit in the classroom with her headset on and listen both to the text and to the lecturer, without anyone thinking that this is strange or even noticing it. Thus the stigmatization problem is minimized at the same time allowing the user to attain more self-confidence in academic arena.

In Iris's and Josh's case, the results of the experiment we conducted have convinced the teacher and given her the will and the encouragement she needed in order to support this kind of iPad use in her class. Silently and inconspicuously for the time being.

Finally, future work as we see it lies in conducting series of smaller studies showcasing how different impairments may be helped, followed by larger studies validating the smaller ones and convincing the school leadership and the policymakers as well.

### ACKNOWLEDGMENT

The authors are grateful to Akademika, the University bookstore and publishing house for providing the iPads for the study, Gyldendahl academic publishing house for providing digitalized text the children could use and the University Library for the support.

### REFERENCES

- [1] A. Gasparini and A. L. Culén, "Children's Journey with iPads in the Classroom," presented at the Opportunities and Challenges when Designing and Developing with Kids @ School at the Interaction Design for Children Conference (IDC 2011), Ann Arbor, Michigan, 2011.
- [2] A. L. Culén and A. Gasparini, "iPad: a new classroom technology? A report from two pilot studies", INFUTURE Proceedings, 2011, pp. 199-208.
- [3] 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.
- [4] B. A. Nardi and V. L. O'Day, Information ecologies : using technology with heart. Cambridge, Mass.: MIT Press, 1999.
- [5] S. Turkle, The second self : computers and the human spirit. New York: Simon and Schuster, 1984.
- [6] N. Shah, "Special Ed. Pupils Find Learning Tool in iPad Applications," Education Week, vol. 30, 2001, pp. 1-1, 16, 17.
- [7] N. Humphrey, "Facilitating a positive sense of self in pupils with dyslexia: the role of teachers and peers," Support for Learning, vol. 18, 2003, pp. 130-136.
- [8] T. D. Lackaye and M. Margalit, "Comparisons of Achievement, Effort, and Self-Perceptions Among Students With Learning Disabilities and Their Peers From Different Achievement Groups," Journal of Learning Disabilities, vol. 39, 2006, pp. 432-432-46.

- [9] University of Oslo. "Dysleksi: lese- og skrivevansker", available at: <http://www.uio.no/studier/admin/tilrettelegging/dysleksi/> (accessed September 16, 2011).
- [10] E. A. Draffan, D. G. Evans, and P. Blenkhorn, "Use of assistive technology by students with dyslexia in post-secondary education," *Disability and Rehabilitation: Assistive Technology*, vol. 2, 2007, pp. 105-116.
- [11] N. Dirää, J. Engelen, et al., "The Use of ICT to Support Students with Dyslexia. HCI and Usability for e-Inclusion", *Lecture Notes in Computer Science Vol. 5889*, 2009, pp. 457-462.
- [12] P. Williams, R. H. Jamali, and D. Nicholas, "Using ICT with people with special education needs: what the literature tells us," *Aslib Proceedings*, 2006, vol. 58, pp. 330-345.
- [13] T. Ouimet and E. Balaban, "Auditory stream biasing in children with reading impairments," *Dyslexia*, vol. 16, 2010, pp. 45-65.
- [14] H. Sampath, J. Sivaswamy, and B. Indurkha, "Assistive systems for children with dyslexia and autism," *SIGACCESS Access. Comput.*, 2010, pp. 32-36.
- [15] K. N. Deibel, "Understanding and Supporting the Adoption of Assistive Technologies by Adults with Reading Disabilities," *University of Washington Ph.D.*, University of Washington, United States -- Washington, 2011.
- [16] B. Flyvbjerg, B. Flyvbjerg, "Case Study," in *The SAGE handbook of qualitative research*, N. K. Denzin and Y. S. Lincoln, Eds., ed Los Angeles: Sage, 2011, pp. XVI, 766 s., ill.
- [17] L. C. Larson, "Digital Readers: The Next Chapter in E-Book Reading and Response," *The Reading Teacher*, vol. 64, 2010, pp. 15-22.
- [18] B. A. O'Brien, J. S. Mansfield, and G. E. Legge, "The effect of print size on reading speed in dyslexia," *Journal of Research in Reading*, vol. 28, 2005, pp. 332-349.

## Spatial Ability and Map-Based Software Applications

M. L. Rusch<sup>a</sup>, S. M. Nusser<sup>b</sup>, L. L. Miller<sup>c</sup>, G. I. Batinov<sup>c</sup>, K. C. Whitney<sup>c</sup>

<sup>a</sup>*Department of Neurology and Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, IA 52242*

<sup>b</sup>*Department of Statistics, Iowa State University, Ames, IA 50011-1210*

<sup>c</sup>*Department of Computer Science, Iowa State University, Ames, IA 50011-1041*

### ABSTRACT

Location-based applications are growing in importance as agencies are placing more and more computing into their field applications. The development of software for these applications needs to consider the wide range of user skills. The present work looks at the impact of spatial ability on a typical Census Bureau application (address verification). A study of a text guided software system for address verification was conducted. The participants were tested to determine their logical reasoning, visualization, and perspective taking abilities. The participants performed a set of address verification tasks using a tablet in a stationary environment. The study and results are presented and discussed.

Keywords: usability, spatial ability.

### I. INTRODUCTION

Human computer interaction (HCI) researchers have recognized the importance of individual differences in cognitive abilities for designing effective software applications. Zhang and Norman [15] argue that a cognitive task is never solely dependent upon the internal mindset of the users nor is it solely related to effectiveness of software design. Both individual differences and system design have the potential to influence computer performance. Among these differences, spatial ability has been found to be one of the strongest predictors of human computer performance [2,3,4,12]. Lohman [8] defined spatial ability as “the ability to generate, retain, retrieve, and transform well-structured visual images” (p. 98).

Spatial ability has several dimensions. Two important components are visualization and orientation. Spatial visualization has been defined as the “ability to manipulate or transform the image of spatial patterns into other arrangements” [5, p. 173]. Spatial orientation has been defined as the “ability to perceive spatial patterns or to maintain orientation with respect to objects in space” [5, p. 149]. Visualization and orientation have been shown to be distinct from one another [6] yet most research on computer performance examines either one or the other [9]. The importance of spatial ability in software user performance suggests that it may be advantageous to the user if software systems were able to accommodate individual differences in spatial ability. Sein, Olfman, Bostrom, and Davis [10] investigated visualization ability in relationship to the usage of three applications (email, modeling software, and operating systems). They found that persons with high visualization skills learned fastest on all of the applications.

Another area related to interface design that we found support for in the literature is guided systems. Berger et al. [1] found that student’s performance in discovery systems was dependent on the level of their cognitive abilities. Sein et al. [10] showed that low spatial users’ performance could be improved by reducing the need for discovery. Meanwhile, Zhang [14] showed that externalization of help aids and extra navigation information can offset the costs of using them. Vicente et al. [11] suggest the use of a task list and instructions for users as a possible accommodating strategy for users of a spatial user interface.

To investigate these ideas, we developed an address verification software system. In this task, Census Bureau field staff evaluate whether the location of an address on the ground is properly represented on the map. If a map error is detected, the location of the housing unit on the map is modified to correct the error. This task can be cognitively demanding. In the field, the user makes a comparison between the location of a housing unit as represented on the map and on the ground.



Figure 1: The computer set up used in the experiment.

Our goal was to investigate whether spatial ability (especially visualization and perspective taking) plays a role in this task and whether the use of task lists and relevant instructions would help bridge performance gaps between persons with low and high spatial ability by reducing the need for discovery. We tested whether using a guided system would reduce the need for discovery for participants with lower spatial ability. We hypothesized that spatial ability would impact the map operations (e.g., zoom and pan) as well as the overall performance (i.e., time and accuracy).

## II. METHOD

### A. Overview

We conducted a user study to evaluate whether spatial ability affected user performance for an address verification task using software interfaces with or without textual protocol and software guidance. We recruited 24 subjects from the community to perform 10 address verification scenarios. Subjects were grouped by gender and age. Within each group, each subject was randomly assigned to the guided or unguided interface treatment. Unlike our paper map study [13], the

present experiment was designed to impose a rigid protocol on the participants. The application recorded the time it took participants to perform each step in the procedure, the number of attempts to match each address, the number of attempts to fix the map, the accuracy in fixing the map, and the number of times specific buttons or other software tools were used.



a) Guided User Interface.



b) Unguided User Interface.

Figure 2. The guided and unguided interfaces.



## B. Participants

The twenty-four subjects were recruited through fliers posted on the Iowa State University campus, local grocery stores, coffee shops, the public library, and word of mouth. This method was used because it mimicked recruiting strategies used by the Census Bureau to recruit address listing staff. The twenty-four participants were equally split by age ( $\leq 30$ ,  $> 30$ ) and by gender (female, male).

## C. Experimental Task and Computing Environment

The experimental task involved comparing a housing unit configuration on the ground (simulated with photographs of the two sides of the street – Figure 1) with the corresponding information in the map. Possible outcomes from comparing ground and map locations are: 1) the ground situation is correctly reflected in the map requiring no further action; 2) the map has an error of commission that requires a map spot to be removed; 3) the map has an error of omission that requires a map spot to be inserted; and 4) the map has an error in the housing unit location that requires the map spot to be relocated.

To successfully perform the task, the following steps need to be executed: 1) find the address on the ground (i.e., in the photos presented to the subject), 2) locate the address on the software map, 3) answer a question posed by the software as to whether or not the address was on the map, 4) if so, answer a question posed by the software as to whether or not the address was in the correct location on the map, and 5) fix the map if an error was identified. Software was developed to instantiate the experimental task. The software generated displays of photographic images of the ground setting on two monitors, one for each side of the street (Figure 1). In addition, the software presented a map-based interface on a tablet PC. The two monitors used in the second session to display the street photographs were Dell UltraSharp 2000FP 20-inch Flat Panel Monitors (16 inches in width and 12 inches in height). The physical dimensions of the map software on the tablet PC were reduced to emulate the size of a handheld. The specific measurements were 2 1/4 inches in width by 3 inches in height for the active interface area and 2 1/16 inches in width and 1 7/8 inches in height for the map display area. The computer used to display the interface was a Gateway Tablet PC M1300. This tablet had a 12.1-inch active matrix LCD color screen and was configured in a landscape display for the experiment (9 3/4 in x 7 1/4 in) (Figure 1). The interface mimicked the size of a handheld computer that might

be used in the field. The guided version of the software displayed an interface that included guidance on what the user should be doing. The unguided version of the software had the same functionality and layout, but provided no guidance (Figure 2).

The guided version included a yellow box at the top of the screen that provided real-time feedback on the step to be executed by the subject. To the left was a list of steps that the subject had to accomplish to complete each scenario. As the user progressed through each step, the current step was highlighted within the list. To the right was an instruction box that provided information about what actions needed to be accomplished on each screen to complete the step. For example, if the user was on a screen in which they were required to fix the map, the screen would tell them one of the specific fixes that needed to be accomplished, such as “Tap delete button”. Map-related functions were the same on both interfaces, and included zoom, pan, reset map, add map spot, and delete map spot. Both interfaces also included an address bar that presented the target address for each of the 10 different scenarios. The software recorded each user action and generated a summary of performance measures for analysis. Specific variables included time spent on each screen, number of attempts to answer each address matching question, positional accuracy in fixing maps, and number of times each map tool was used.

For the photographic images, we used manipulated photos of streetscapes. The original photos were taken in areas of Story County that were not highly trafficked so that subjects would not recognize street configurations. The experiment used maps that were compiled based on Iowa data from Black Hawk County and the Department of Transportation (DOT). These maps were similar to TIGER/Line shape files that are used by the Census Bureau. The manipulation created settings that challenged the users in ways that were consistent with the objectives of the study. For example, we removed a structure from a photo to create a vacant lot on the ground where the map included an existing map spot. In developing the scenarios, we created variation in relation to six factors. These factors included photo, street name, road configuration (e.g., four-way intersection, three-way intersection, etc.), rotation (e.g., north up, south up, etc.), map, and corrective action required.

## D. Experimental Procedure

The experiment involved two sessions with subjects. The first session was used to test the subjects and the subjects performed the map task in

the second session. During the first session, each subject was presented with an informed consent form. After having read this form and signed it, cognitive tests were administered to the subjects. A test script was used to ensure consistency. Three cognitive tests were administered to each subject. These tests included Ekstrom et al. (1976) paper-based assessments on visualization (VZ-2) and logical reasoning (Ekstrom et al. 1976) and the Kozhevnikov et al. (2006) computerized perspective taking assessment on orientation. The Inference Test on reasoning was administered first and the Paper-folding Test on spatial visualization was administered second. After the paper-based tests were completed, the subjects were taken to a computer lab where they completed the background questionnaire. Next they were trained on the Perspective Taking software (PT) and then they proceeded to complete the test. The PT results were compiled by a research team member who was not involved in working with subjects and used to randomize subjects to guided and unguided treatments within age-gender groups. The randomization procedure ensured balance in spatial ability across treatments within these groups.

The second sessions took place throughout the two weeks that followed the first session and lasted approximately one hour each. When a subject returned, s/he was informed that s/he would perform a task that comparing the location of a target housing unit on the ground with its representation on the map. Subjects were trained on the task procedure using an example scenario that was based on two color paper printouts. One color printout included two street photos and the other included a zoomed-in map that emulated the map that would be displayed on the software interface. The tablet touch screen calibration was performed by each subject to ensure that the tablet was sensitive to the user's handedness and the way in which s/he used the stylus. Finally, the user was trained to use the software to accomplish the experimental task and allowed to practice with two computerized practice scenarios. After training, the subject proceeded to complete each of 10 test scenarios. After completing the experiment, the subject received a \$30 gift card.

### E. Analysis Methods

The impact of interface treatment (i.e., guided or unguided) and associations with demographic and cognitive ability covariates on the subjects' behavioral and performance measures were evaluated using regression procedures. Response variables included the time required to perform each scenario, the accuracy of locations for addresses that required adding or moving map spots, the number of

times the pan button was used, and the number of times the zoom button was used. Accuracy of a newly placed housing unit map spot was derived by computing the distance (in meters) between the centroid of the parcel in which the housing unit was located and the location of the housing unit inserted by the subject. Because preliminary analyses indicated that the location accuracy variable required a transformation to meet regression analysis assumptions, a log transformation was applied to this variable prior to fitting the regression model. The interface treatment variable was expressed as an indicator variable indicating whether the subject was assigned to the guided treatment or not. Demographic variables (expressed as classification variables) included in the model were age category (18-29 years of age, 30-39 years of age, 40-59 years of age, or 60 years and older) and gender. For the cognitive tests, we standardized for visualization, perspective taking, and logical reasoning. To avoid problems with collinearity among spatial ability measures, we ran 3 sets of analyses: a) we used only VZ, b) both VZ and the *spatial difference* (VZ-PT) were used, c) average (VZ+PT)/2 and the spatial difference were used. Regression models were fit using an ordinary least squares (PROC GLM in SAS, citation). We examined residuals for departures from assumptions of homogenous variance and linearity. Tests of whether regression parameters were equal to zero were conducted to identify which covariates were associated with each response variable.

### III. RESULTS

Table 1 presents the test results from the analysis for time, log accuracy, zoom button usage, pan button usage, and map reset button usage. Guidance was not related to any performance measure, after accounting for the other explanatory variables, except gender. In analysis c) we found a significant negative association between average of visualization and perspective taking and time to perform the task. The estimated regression coefficient was -320 (SE=145) indicating that as the average of the visualization and perspective taking standardized test scores increased by one unit, the average time spent on completing the full exercise was reduced by an estimated 320 seconds (holding other variables constant). Analyses a) -327 (SE=189) and b) -245 (SE=74) found similar results for association between VZ and the time required.

There was a significant negative association in analysis c) between error in housing unit location (log meters) and spatial difference. The estimated

**Table 2. P-values for ANOVA F-tests for each interesting performance variable.**

Source	Time (sec)	Accuracy (log m)	Zoom (# zoom actions)	User Map Resets (# reset actions)	Pan (# pan actions)
Age			0.02 <sup>b</sup>		
Gender		0.02 <sup>c</sup>	0.009 <sup>b</sup>		
Gender * Interface			0.01 <sup>b</sup>		
VZ	0.001 <sup>b</sup>		0.03 <sup>b</sup>	0.03 <sup>a</sup>	0.03 <sup>a</sup>
Spatial Difference (VZ-PT)	0.006 <sup>b</sup>	0.02 <sup>c</sup>			0.02 <sup>c</sup>
Spatial Average	0.05 <sup>c</sup>				

- a) Analysis only with VZ.
- b) Analysis with both VZ and Spatial Difference.
- c) Analysis with both the average and Spatial Difference.

regression coefficient was -.69 (SE=.26), indicating that for every unit increase in the difference between visualization and perspective taking standardized test scores, the user-determined housing unit locations was an estimated .69 log meters closer to the target location. There was also a significant association between gender and accuracy in analysis c). The estimated regression coefficient was 1.25 (SE=.48) which indicated that females tended to be 1.25 log meters less accurate than males.

Analysis c) found significant negative association between age and the use of the zoom. The estimated regression coefficient was -5.18 (SE=2.15), which meant that older subjects tended to make less use of the zoom tool. Analysis b) saw a negative association -5.56 (SE=2.36) between VZ and zoom. A positive association 19.82 (SE=6.83) also showed up between guided females and the use of zoom in analysis b). A significant negative association of the differences between visualization and perspective taking scores and use of the pan buttons was found in analysis c). The estimated regression coefficient was -52.13 (SE=19.94) which meant that subjects with high visualization relative to perspective taking scores tended to make less use of the pan tool. Finally, a negative association was found in analysis a) for the use of the reset map button (-1.93 (SE=0.82)).

#### IV. DISCUSSION

An important goal of this research was to investigate the relationship of spatial ability and user software performance for a map-related task in

relation to two specific sub-factors of spatial visualization and orientation (i.e., perspective taking) abilities. It is clear from the analyses that the results were sensitive to the relationship between the two spatial parameters used. To get a complete understanding, we used the three combinations of parameters (a,b,c) shown in the legend of Table 1. Our results indicate the maps were more sensitive to VZ than the average of the two spatial parameters. We found that higher visualization scores tended to be correlated with faster performance times and fewer map operations (zooms, pans, and map resets). The association between spatial ability and user performance is consistent with findings from a large body of literature in software use (Dahlbäck et al., 1996; Egan, 1988; Vicente, Hayes, & Williges, 1987; Egan & Gomez, 1985). In addition, our results extend this finding to map-based interfaces.

We also saw that for this set of data that there were differential effects of spatial ability sub-factors corresponding to visualization and perspective taking. Subjects with higher spatial differences were able to more accurately record the location of addresses that were missing from the map.

Pan usage was also lower for subjects with higher spatial differences, which is a likely result of pan usages being lower for subjects with higher VZ scores. Older subjects tended to use the zoom tool less frequently. Based on their successful completion of the tasks, there isn't any indication that this impacted their overall performance.

Spatial ability has also been found to vary by gender, and when this factor is significant, results

indicate that spatial ability tends to be higher for men relative to women (Linn & Peterson, 1985). We found that on average, male subjects most accurately placed housing units on the map when the ground situation showed a housing unit that was not initially present on the map. In addition females relied to some degree on the guided interface.

The logical reasoning abilities of the participants were not significant for any of the performance parameters.

The lack of a relationship between the availability of the software layout (i.e., guided treatment) and spatial ability was somewhat surprising. Based on the connection between discovery of the software structure and visualization scores in the literature, one would have expected more value from the guided interface. The question of interest is whether the fact that our software broke the task into a series of rather simple self-contained subtasks (Figure 2) reduced the participants' need for discovery in the sense detailed by Sein et al. (1993). There are two directions we will be able to go to better understand why we didn't see a relationship. One issue is the complexity of the task. We are currently conducting two studies to provide more information on what participant skills are being used in the address verification task. Future experiments will be designed to look at other forms of guidance to help us determine whether our guidance structure and/or content was inadequate.

#### ACKNOWLEDGEMENTS

We gratefully acknowledge the scientific guidance of Elizabeth Murphy and Kathleen Ashenfelter of the U.S. Census Bureau. This research was funded in part by a professional services agreement with the U.S. Census Bureau and by the National Science Foundation grant No. 0092696.

#### REFERENCES

[1] Berger, C.F., Lu, C.R., Belzer, S.J., & Voss, B.E. (1994). Research on the use of technology in science education. In Gabel, D. L. (Ed.), *Handbook of Research on Science Teaching and Learning*, Simon and Schuster, New York, pp. 466–490.

[2] Dahlbäck, N., Höök, K., & Sjölander, M. (1996). Spatial Cognition in the Mind and in the World: The case of hypermedia navigation. *The Eighteenth Annual Meeting of the Cognitive Science Society, CogSci '96*, University of California, San Diego, July.

[3] Egan, D. (1988). Individual differences in human-computer interaction. In: M. Helander (ed.),

*Handbook of Human-Computer Interaction*, (pp. 543-568). Amsterdam: Elsevier Science Publishers.

[4] Egan, D.E., & Gomez, L.M. (1985). Assaying, isolating and accommodating individual differences in learning a complex skill. In R. Dillon (Ed.), *Individual differences in cognition*. New York: Academic Press.

[5] Ekstrom, R.B., French, J.W., Harman, H.H., & Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.

[6] Kozhevnikov, M., & Hegarty, M., (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory and Cognition*, 29, 745-756.

[7] Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56, 1479-1498.

[8] 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.

[9] Pak, R., Rogers, W. A., & Fisk, A. D. (2006). Spatial ability subfactors and their influences on a computer-based information search task. *Human Factors*, 48, 154-165.

[10] Sein, M., Olfmann, L., Bostrom, R., & Davis, S. (1993). Visualization ability as a predictor of user learning success. *International Journal of Man – Machine Studies*, 39, 599 – 620.

[11] Vicente, K.J., and Williges, R.C., (1988). Accommodating individual differences in searching a hierarchical file system, *International Journal of man Machine Studies*, 29, 647 – 668.

[12] Vicente, K., Hayes, B.C., & Williges, R.C. (1987). Assaying and isolating individual differences in searching a hierarchical file system. *Human Factors*, 29, 349-359.

[13] Whitney, K.C., G.J. Batinov, L.L. Miller, S.M. Nusser, and K.T. Ashenfelter. (2011). Exploring a Map Survey Task's Sensitivity to Cognitive Ability. *The Fourth International Conference on Advances in Computer-Human Interactions*. Gosier, Guadeloupe, France. 63-68.

[14] Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, 21, 2, 179-217.

[15] Zhang, J., & Norman, D.A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, 18, 87-122

# A Presentation Support System by Expanding Embodiment with a Mobile Touchscreen Device

Michiya Yamamoto  
 Taku Murabayashi  
 School of Science and Technology  
 Kwansei Gakuin University  
 Sanda, Hyogo, Japan  
 michiya.yamamoto@kwansei.ac.jp  
 asn48941@kwansei.ac.jp

Tomio Watanabe  
 Faculty of Computer Science and Systems Engineering  
 Okayama Prefectural University  
 Soja, Okayama, Japan  
 watanabe@cse.oka-pu.ac.jp

**Abstract**— Most of the people make presentations by using tools such as pointers. However, the use of information devices such as laser pointers and mice can restrict the presenter’s embodied motions and actions. In this paper, we propose a presentation support system that can expand embodiment by using a portable touch screen device. First, we develop a prototype of the system by using an iPhone, and then, we perform experimental evaluation. Next, we perform the evaluation of the modes of PPTouch, and the effectiveness of the system in various situations is clarified. Then we perform comparison with conventional pointers, and evaluation by participants. Both experiments also show the effectiveness of PPTouch.

**Keywords**- presentation support; expansion of embodiment

## I. INTRODUCTION

In presentations, tools such as pointers are used together with embodied motions and actions of body language and gestures [1]. However, the use of information devices such as laser pointers and mice can restrict the presenter’s embodied motions and actions. For example, when using a pointer, a presenter can explain the slides with various motions and actions depending on the content. Nevertheless, when the presenter uses a laser pointer or mouse, the participants may have difficulty seeing the presenter’s movements, and this makes the presentation harder to understand.

Several researchers have studied this problem and attempting to provide possible solutions by using information devices. For instance, Murata et al. developed a presentation tool that superimposes a shadow, as in an OHP presentation [2]. Shimizu et al. developed a system that enables multiple pointing using mobile devices [3]. Some applications are available on App Store for iPhone [4, 5] or iPhone software [6]. However, these researches are based on functional support.

On the other hand, we have developed a presentation support system that can share embodied rhythms between the

presenter and the participants and demonstrated the effectiveness of our system. Our speech-driven embodied entrainment presentation support system can generate a listeners’ nodding reaction via visual and auditory information and create a sense of unity between presenters and participants [7]. We have also introduced the presenter’s embodied motions and actions into the system via a pen display and enabled the rhythmic communication of information [8]. In this study, we introduce the concept of PPTouch. With PPTouch, presenters can use mobile devices with embodied motions and actions as if they are a part of their body by expanding embodiment. We also develop a prototype and perform evaluation experiments.

## II. CONCEPT

Figure 1 shows the concept of PPTouch. The presenter moves the finger-shaped cursor on the screen by operating a mobile touchscreen device, and the presenter can feel as if the embodiment is expanded.

The mobile device has a touch screen and is WiFi-enabled. Its use allows the presenter to gesture from various positions in various situations. Moreover, because the touchscreen has an intuitive interface, it can be used as if it is a part of the presenter’s body.

In addition, the embodied visual effects such as the finger-shaped cursor enhance the efficiency of communication.

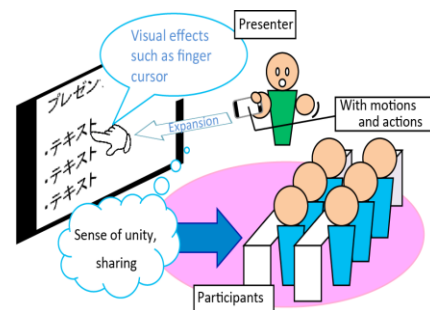


Fig. 1 Concept of PPTouch.

### III. PPTOUCH

#### A. System Configuration

Figure 2 shows the system configuration of PPTouch. For the mobile touchscreen device, we used an Apple iPhone (with iOS 4.2.1) and developed the software using XCode 3.2.6. When a presenter touches the iPhone screen, the software detects the input. When cursor operations are input, messages are sent to the PC (HP, EliteBook 8730w) using UDP. When page-control or effect-control commands are input, messages are sent using TCP. The PC handles the received messages and controls the presentation software (Microsoft, PowerPoint 2010). A finger-shaped cursor is projected on the main screen and controlled using the iPhone.

#### B. Modes for Expanding Embodiment

On the iPhone, we display a copy of the projected slide in the upper section of the touchscreen. The lower section contains the buttons that control the visual effects. We have developed three operation modes (see Figure 3):

##### Mode A: Finger cursor

A finger-shaped cursor is displayed on the iPhone screen and synchronized with the cursor on the main screen. By moving the iPhone cursor, presenters can point to the same point of the main screen.

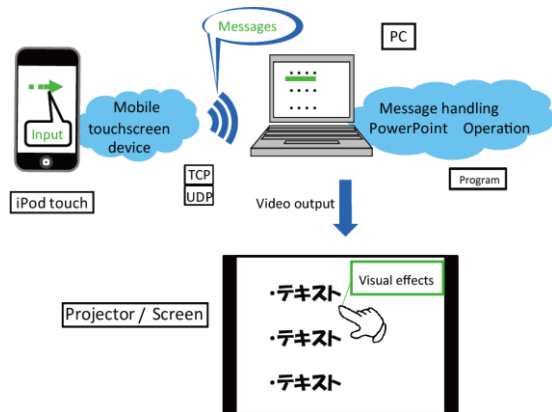


Fig. 2 System configuration.

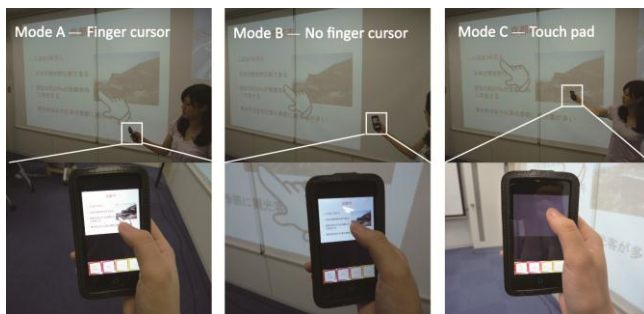


Fig. 3 Three modes for expanding embodiment.

##### Mode B: No finger cursor

Presenters can touch a point on the iPhone slide, and the cursor on the main screen will move to that point. Presenters feel as if they are directly touching the main screen. No finger-shaped cursor is displayed.

##### Mode C: Touchpad

As for a PC touchpad, the finger cursor on the screen moves according to the distance to which it is dragged. There is no slide or finger cursor on the mobile device.

### IV. COMPARISON BETWEEN MODES

#### A. Method

To compare the three modes, we assumed the four presentation scenarios shown in Table 1 and conducted an experimental evaluation.

First, we explained PPTouch and the experimental scenario and asked the subjects to become familiar with the PPTouch system. We then selected a scenario, randomly. We evaluated the modes via a paired comparison ( $3P_2 = 6$ ) by asking the subjects to give a presentation using the three modes in random order. For the presentation, we prepared three slides that the subjects could easily explain without any previous knowledge. The slides included red keywords, as shown in Figure 4 and the subjects were asked to emphasize these during the presentation. We followed this procedure for four scenarios.

After that, the subjects were instructed to perform seven-point bipolar rating scale from 1 (not at all) to 7 (extremely), over all evaluation, and free comments of each mode as an overall evaluation. The experiment was performed by 15 pairs of 30 Japanese students. Figure 5 shows the example scene of the experiment.

Table 1. Scenarios.

Scenario	Details
Large screen	Presenter can not touch the screen.
Medium screen	Presentation in a room with a medium screen. Presenter can use a pointer.
Speech	Presenter looks at the audience.
Meeting	Presenter and participant sit side by side.

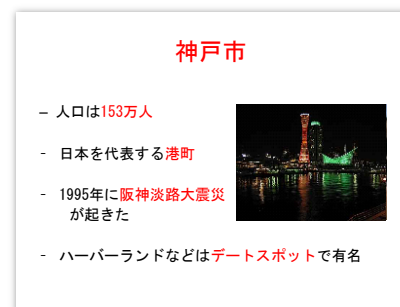


Fig. 4 Example slide.

B. Results

Table 2 shows the results of paired comparison. Here, the Bradley-Terry model [9] was fitted to the result for a quantitative analysis, and preference  $\pi$  was estimated. Figure 6 shows the result. As a result, in a large screen scenery the “mode A - finger cursor” and “Mode B - no finger cursor” were rated about twice than “Mode C - touchpad”. In a medium screen scenario, mode B was rated more than twice as high as mode C. In speech, mode A was rated extremely high. In meetings, there was no large difference.

Fig. 7 shows the result of overall evaluations. Mode B was rated highest. Mode A followed to it, however, mode C was not selected so much.

Figure 8 shows the results of seven-points bipolar rating. The result of average, SD, and the Friedman test are also shown in this figure. For “I felt as if I was touching the screen” and “I looked at the device frequently”, there was a significant difference at a significance level of 1%; mode A and mode B were rated higher than mode C. There was also significant difference in “I looked at the main screen”.

Table 3 shows the result of users’ comments. Mode A and mode B were commented favorably, however, some of the subject pointed out that “I had to check the finger position”, etc.

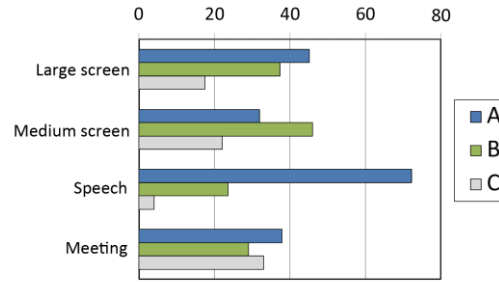


Fig. 6 Preference  $\pi$ .

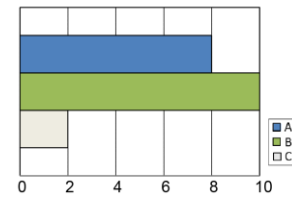


Fig. 7 Overall evaluations.

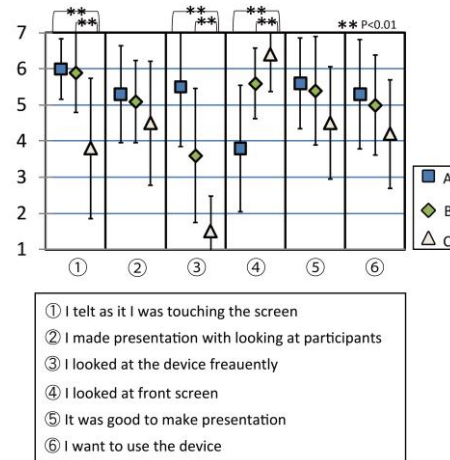


Fig. 8 Results of seven-point bipolar rating.

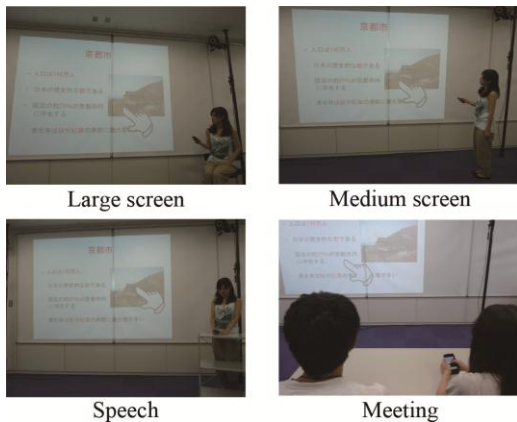


Fig. 5 Four experimental sceneries.

Table 2. Results of paired comparison.

Large screen	A	B	C	total
	A	15	23	38
B	15	19	34	
C	7	11	18	

Medium screen	A	B	C	total
	A	12	18	30
B	18	20	38	
C	12	10	22	

Speech	A	B	C	total
	A	22	29	51
B	8	25	33	
C	1	5	6	

Meeting	A	B	C	total
	A	16	17	33
B	14	13	27	
C	13	17	30	

Table 3. Users’ comments.

Positive comments	
A	The cursor made it easy to know where was pointing.
B	I could operate intuitively. I could use it unconsciously.
C	I could perform familiar operations in a standing position.
Negative comments	
A	I had to look for the cursor first.
B	I had to check the finger position.
C	I had to check the pointer's position constantly.

C. Discussion

Mode B (no finger cursor) received the highest evaluations except in the speech scenario. Thus, mode B was confirmed to be useful in various situations. Mode A (finger cursor) was also rated highly, especially in the speech scenario. Thus, a presenter can use this device even when facing the audience. Modes A and B were rated highly overall and in the category *I felt as if I was touching the screen*. Thus, our device is useful for presentation support. On the other hand, mode C received some positive evaluations; this was because some subjects favored a familiar operation on a mobile device. These results show the effectiveness of PPTouch.

V. COMPARISON WITH CONVENTIONAL POINTERS

A. Method

Next, we compared the use of a pointer, a laser pointer, and PPTouch in three scenarios: large screen, medium screen, and meeting (see Figure 9). In the experiment, we asked each pair of subjects to choose the appropriate mode and become familiar with PPTouch. Then, we asked one person in each pair to identify himself/herself as the presenter and the other as the participant.

We selected one scenario randomly. We evaluated the modes via a paired comparison ( ${}_3P_2 = 6$ ) by asking the subjects to give a presentation using the three modes in random order. Then, the subjects were instructed to perform seven-point bipolar rating scale from 1 (not at all) to 7 (extremely), and free comments of each mode. The experiment was performed by 15 pairs of 30 Japanese students.

B. Results

Table 4 shows the results of paired comparison. (Pointer is described as S, Laser pointer is L, and PPTouch is P hereafter). Here, the Bradley-Terry model was fitted to the result for a quantitative analysis, and preference  $\pi$  was

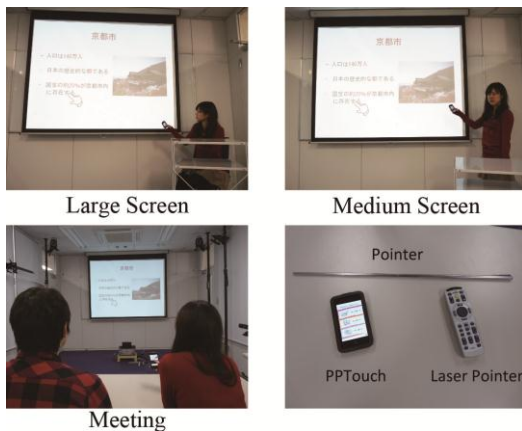


Fig. 9 Three devices and three sceneries.

estimated. Figure 10 shows the result. As a result, laser pointer was rated high in meeting. PPTouch was rated high in all scenery. Figure 11 shows the results of seven-point bipolar rating. The result of average, SD, and the Friedman test is also shown in this figure. Friedman’s test revealed significant differences of 1% between the pointer and PPTouch for all items. There was a 1% significant difference for “I was good to make the presentation”, “My pointing was clear”, and “I want to use the device” between the pointer and laser pointer. There was a 5% significant difference for “I could move or gesture freely”. On the other hand, there was no significant difference between the results for the laser pointer and PPTouch.

We then classified the results on the basis of the response to “I want to use the device”. Figure 12 shows the results. Seventeen subjects preferred PPTouch, and thirteen preferred other devices. As shown in Figure 12, seventeen rated PPTouch highly. Friedman’s test revealed significant differences of 1% or 5% between the pointers for all the responses. There was a 1% significant difference for “I want to use the device”. There was a 5% significant difference for “It was good to give the presentation”, “I looked at the participants”, and “I felt as if I was touching the screen”. On the other hand, the remaining thirteen participants reported no significant difference between the results for the laser pointer and PPTouch. Thus, PPTouch was rated as highly as the laser pointer.

Figure 13 shows the results by participants. There are significant differences between pointer and other devices at 1% or 5% of significant level. However, there was no significant difference between laser pointer and PPTouch.

Table 5 shows the result of users’ comments. PPTouch commented favorably for both presenter and participants. However, some of the subject pointed out that “Some of the words may be hidden by the finger cursor”, etc.

Table 4. Results of paired comparison.

Large screen	S	L	P	total	Medium screen	S	L	P	total
	S	\	3	0		3	S	\	12
L	27	\	7	34	L	18	\	11	29
P	30	23	\	53	P	20	19	\	39

Meeting	S	L	P	total
	S	\	0	0
L	30	\	20	50
P	30	10	\	40

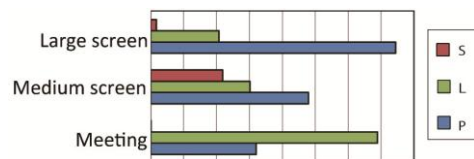


Fig. 10 Preference  $\pi$ .



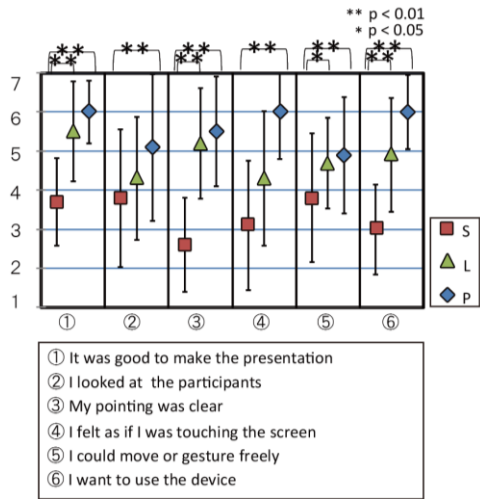


Fig. 11 Results of seven-point bipolar rating (presenters).

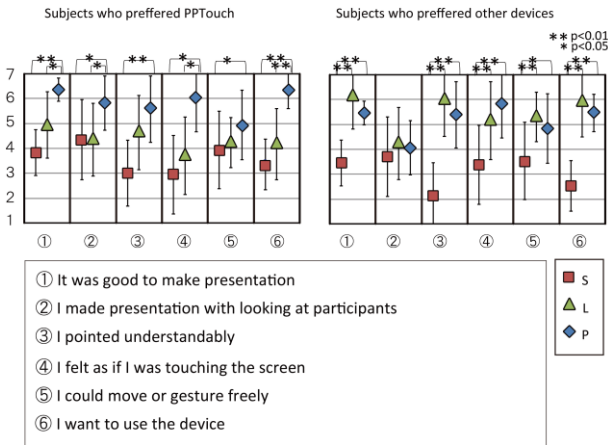


Fig. 12 Results of seven-point bipolar rating (presenters) subdivided by preference.

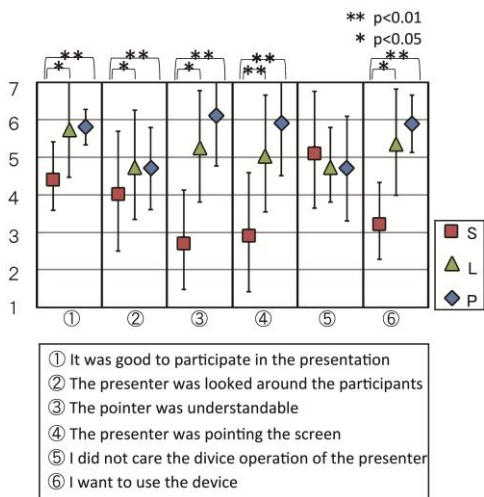


Fig. 13 Results of seven-point bipolar rating (participants).

Table 5. Users' comments.

Presenter	
Positive comments	<ul style="list-style-type: none"> <li>• PPTouch was versatile sufficiently.</li> <li>• The pointed area was easy to understand.</li> <li>• I could make the main presentation without looking at the front screen.</li> </ul>
Negative comments	<ul style="list-style-type: none"> <li>• I had to look at the device till I got used to it.</li> <li>• Some words may be hidden by the finger cursor.</li> <li>• The font size on the screen was small.</li> </ul>
Participants	
Positive comments	<ul style="list-style-type: none"> <li>• Important points were emphasized adequately.</li> <li>• The pointing was clear.</li> </ul>
Negative comments	<ul style="list-style-type: none"> <li>• Some of the words may be hidden by the finger cursor.</li> <li>• The presenter looked at the device during the presentation.</li> </ul>

C. Discussion

The pointer was rated highly in the medium-screen scenario, but its use is limited by its length. The laser pointer was rated highly in the meeting scenario because it was useful when the presenter pointed to the screen from the front. PPTouch was rated highly in all scenarios. Thus, we conclude that PPTouch is sufficiently versatile for various presentation scenarios.

VI. EVALUATION BY PARTICIPANTS

A. Method

According to Lilian's study, Highlighting, Outlining, the Pointing and Emphasizing are important in presentation [1]. As an evaluation experiment by participants the situation with a big screen where PPTouch can be used easily was prepared (see Figure 14 of left). The experiment was conducted by 3 persons (one presenter and two auditors were in a pair).

First, we prepared the slide for each subject (Figure 15), and asked them practice the presentation by using PPTouch. In the practice, we told the method of pointing according to the contents of the slide, how to move the cursor at the presentation, and how to change the size of a cursor. After that, they performed the presentation of the slide using a laser pointer or PPTouch. Then, the participant filled in the seven-points bipolar rating and free questionnaire of the device. These are determined as one set. Then, we changed the device and asked them to fill out the form again. Then, they changed a presenter as a next set. And they performed the three set in total.

An order of the used slides or a device was in random. The experiment was performed by 10 pairs of 30 Japanese students.



Fig. 14 Big screen scenario for evaluation by the participants.



Fig. 15 Example of the slides.

B. Results

Figure 16 shows the results. Friedman’s test revealed significant differences of 1% or 5% between the pointers for the items ③—⑦. There was a 1% significant difference for “The figure was highlighted” and “Explanation of the graph was understandable” and “The pointing was clear” and “I want to use the device”. There was a 5% significant difference for “The text on the slide was clearly emphasized”.

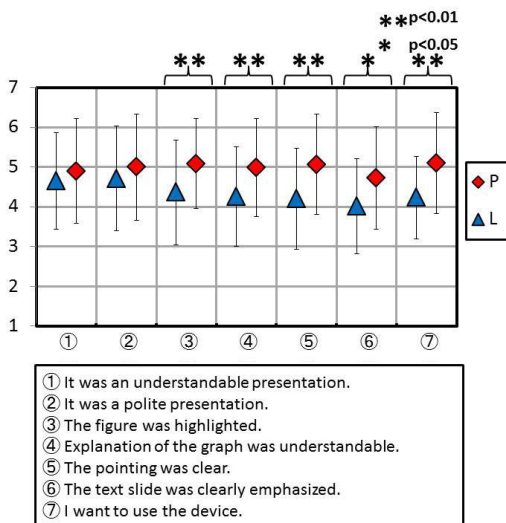


Fig. 16 Results of seven-point bipolar rating (participants).

C. Discussion

Because the cursor position of PPTouch was stable, the participants rated PPTouch higher than laser pointer. As for Emphasizing, it was not rated so high than the other modes. It may be caused by the occlusion by the finger cursor.

VII. CONCLUSION

In this study, we proposed a PPTouch presentation support system that can expand embodiment. We developed a prototype using an iPhone and implemented three operation modes. Then, we performed an experimental evaluation, the results of which showed that PPTouch is useful for presentation support in various scenarios. We then compared PPTouch with conventional pointers and found that PPTouch is sufficiently versatile. Also, we performed evaluation experiment by the participants, and made clear its advantage.

ACKNOWLEDGMENT

This work under our project “Generation and Control Technology of Human-entrained Embodied Media” has been supported by CREST of JST.

REFERENCES

- [1] L. L. Pozzer-Ardenghi and W. Roth, “Gestures: Helping Students to Understand Photographs in Lectures,” in *Connections '03*, 2003, pp. 1-30.
- [2] Y. Murata, K. Kurihara, T. Mochizuki, B. Shizuki and J. Tanaka, “Designing a pen shadow for the presentation tool which superimposes the shadow,” in *IPSJ SIG Technical Report*, 2010-HCI-139, 2010, pp. 1-5. (in Japanese)
- [3] K. Shimizu, T. Tsujino, S. Shiramatsu, T. Ozono and T. Shintani, “Implementing a Presentation Support System Using a Multi-pointing Mechanism,” in *Proceedings of Forum on Information Technology 2009*, Vol.4, 2009, pp. 449-450. (in Japanese)
- [4] i-Clickr PowerPoint Remote: <http://www.senstic.com/iphone/iclickr/iclickr.aspx> <retrieved: Jan., 2012>
- [5] POINTER: <http://www.zentropysoftware.com/ZS/Pointer.html> <retrieved: Jan., 2012>
- [6] S. Boring, D. Baur, A. Butz, and S. Gustafson and P. Baudisch, “Touch Projector: Mobile Interaction through Video,” in *Proceedings of the 28th international conference on Human factors in computing systems*, 2010, pp. 2287-2296.
- [7] H. Nagai, T. Watanabe and M. Yamamoto, “InterPointer: Speech- Driven Embodied Entrainment Pointer System,” in *The 2005 International Conference o Active Media Technology (AMT2005)*, 2005, pp. 213-218.
- [8] Y. Takao, M. Yamamoto and T. Watanabe, “Development of Embodied Visual Effects which Expand the Presentation Motion of Emphasis and Indication,” *LNCS 6762, 2011*, p. 603-612.
- [9] C. Hirotsu, “Analysis of Experimental Data, Beyond Analysis of Variance”, Kyoritsu Shuppan, Tokyo, 1992, pp. 203-217. (in Japanese)

# Designing Mobile Apps for Visually Impaired and Blind Users

## Using touch screen based mobile devices: iPhone/iPad

Javier Sánchez Sierra

Center for Computer Research in Music and Acoustics  
Stanford University  
Palo Alto, CA, USA  
e-mail: jsanchez@ccrma.stanford.edu

Joaquín Selva Roca de Togores

Raylight Soluciones Tecnológicas  
Valencia, Spain  
e-mail: ica@icaudit.org

**Abstract**—Although mobile devices include accessibility features available for visually impaired users, the user interface of the majority of the mobile apps is designed for sighted people. It is clear that “Design for Usability” differs depending if the final user is a sighted user or a visually impaired user. This paper introduces the concept of “Low Vision Mobile App Portal”, which provides a way to access mobile apps specifically designed for visually impaired users. Some design aspects will be described. Preliminary results show some of the low vision controls specifically design for visually impaired users.

**Keywords** - design; mobile App; visually impaired; blind; usability; low vision mobile portal; accesibility; iOS ; iPhone

### I. INTRODUCTION

This research work deals about design, mobile App design centered on low vision users. The concept of *Universal Design* has been widely used in several fields, such as architecture or product design. This term was coined by the Architect Ronald L. Mace [1], and refers to the idea of designing products to be aesthetic and usable by everyone, regardless of their age, ability or status in life.

The most common adjectives used when referring to universal design are: simple, intuitive, equitable, flexibility, perceptible or tolerance for error. The term *Universal Design* is closely related to other terms such as accessibility or usability [2].

With the appearance of the new technologies, the term *accessibility* is extended to *computer accessibility*. The majority of the operative systems include new and innovative solutions for people with disabilities. See Fig.1.

Due the growth of Internet, there is a specific section inside computer accessibility dealing with web accessibility.

Some authors [3-7] have written about this topic, describing assistive technologies for web browsing: speech

recognition, screen magnification or screen reader software.

In 1999, the Web Accessibility Initiative (WAI) published the Web Content Accessibility Guidelines WCAG, to improve the accessibility of the web for people with disabilities.

Since the appearance of mobile devices, human computer interaction has changed significantly, appearing new techniques for usability evaluation [8-9]. The usability tests evaluate the user interface and navigation issues in different environments.

A radical changed has occurred since the development of touch screen based mobile devices, such as the iPhone, iPad or Android devices. In less that a couple of years, gesture based interaction has become a standard on the majority of mobile devices.

It is an emerging area of research since touch displays are more and more present in our everyday life [10-11]. Touch screens provide a great flexibility and a direct access to controls and information, but on the other hand, the physical feedback is lost, making them less accessible to visually impaired and blind users. The goal of our research work is to facilitate low vision users the interaction with devices that use this kind of displays.

So, this paper deals about design. Design of touch based mobile apps usable by visually impaired people. Despite the great effort of hardware manufacturers to include accessibility features in their touch based mobile devices, they are not good enough to obtain a good visually impaired user experience. Since most of the existing apps are designed for sighted users, the accessibility features are not always adequate to obtain a reliable result.

Section II of this paper describes the importance of designing specific mobile apps for low vision users. *Design for Usability* is an essential requirement to achieve a good feedback from visual impaired users when using the apps.

In section III, the *Low Vision Mobile App Portal* is introduced. Thanks to this portal, visually impaired people will have the possibility of accessing an extensive collection of apps specifically designed for visually impaired users.

Later on, section IV describes some of the mobile apps that will be available in the *Low Vision Mobile App Portal*. Traditional apps such as telephone, calendar or contacts need to be reinvented. On the other hand, specific apps for blind users, such as text magnifiers or GPS, need to be effectively designed.



(a)



(b)

Figure 1. Accessibility options: (a) Windows 7 (b) Mac OS X

The paper finishes with some conclusions, which show the benefits of using *Design for Usability* when designing apps for blind and visually impaired users.

II. DESIGN FOR USABILITY FOR BLIND AND VISUALLY IMPAIRED USERS

Following are described some basic concepts needed in order to design specific mobile apps for visually impaired users. iOS devices (iPhone and iPad) are used to illustrate these ideas.

A. User Experience / Usability

One of the best ways to evaluate the effectiveness of a product design is to obtain a good user experience [12-13].

User Experience is about how a person feels about using a product, system or service. User experience highlights the experiential, affective, meaningful and valuable aspects of human-computer interaction [14]. As its name indicates, *user experience* focused on the user.

As example of *user experience* in touch screen based devices, iPhone and iPad have an specific *Human Interface Guidelines* [15], which describes the guidelines and principles that help developers to design a superlative user interface and a user experience for the iOS app. These guidelines are oriented to design apps for sighted users.

Usability describes the quality of user experience.

B. Accesibility in mobile Apps

Accessibility is a general term used to describe the degree to which a product, device, service, or environment is available to as many people as possible [16].

As example, iPhone and iPad include a set of features specifically designed to provide accessibility to users with special needs. Some of these features are: VoiceOver, Voice Control, White on Black, Zoom, Speak auto-text, tactile buttons, giant fonts, hands-free speakerphone, audible, visible and vibrating alerts or assignable ringtones.

In order to create accessible apps, Apple Accessibility Programming Guide [17] helps iOS developers make their applications accessible to low vision users, using the Voice over feature.

Once user experience, usability and accessibility have been defined, we can concentrate on mobile Apps design. The vast majority of the 425.000 iOS apps available at the App Store have been designed for sighted users. The user interface and the usability tests of these apps have been designed focusing on sighted users. If the designer decides to make the app accessible for low vision or blind users, he can add an extra layer with useful labels and hints that will be used by VoiceOver feature.

Figure 2a shows this design situation. A mobile app designed for sighted users with an extra layer including accessibility features. Although blind users could use this mobile app, the user interface has not been conceived for blind users. In this case, design for usability applies only for sighted users and blind user experience is not assured at all.

Figure 2b shows a different design scenario. In this case, the mobile app has been designed centered on low vision people. Design for usability applies directly to visually impaired and blind users. It affects to design aspects of the user interface, such us the size of the controls, position, shape, image contrast, brightness, etc. Voice over accessibility features are not just an extra layer, they are directly connected to the main user controls.

If we want to design apps for visually impaired people, we should go to the scenario shown in Figure 2b, instead of adding accessibility features to apps that are not initially conceived for visually impaired users.

According to this premise, specific apps are needed for visually impaired users in order to obtain the best user experience. Even for those general apps, such us calendar, contacts or phone, a specific design app will increase dramatically the usability and user experience for visually impaired users.

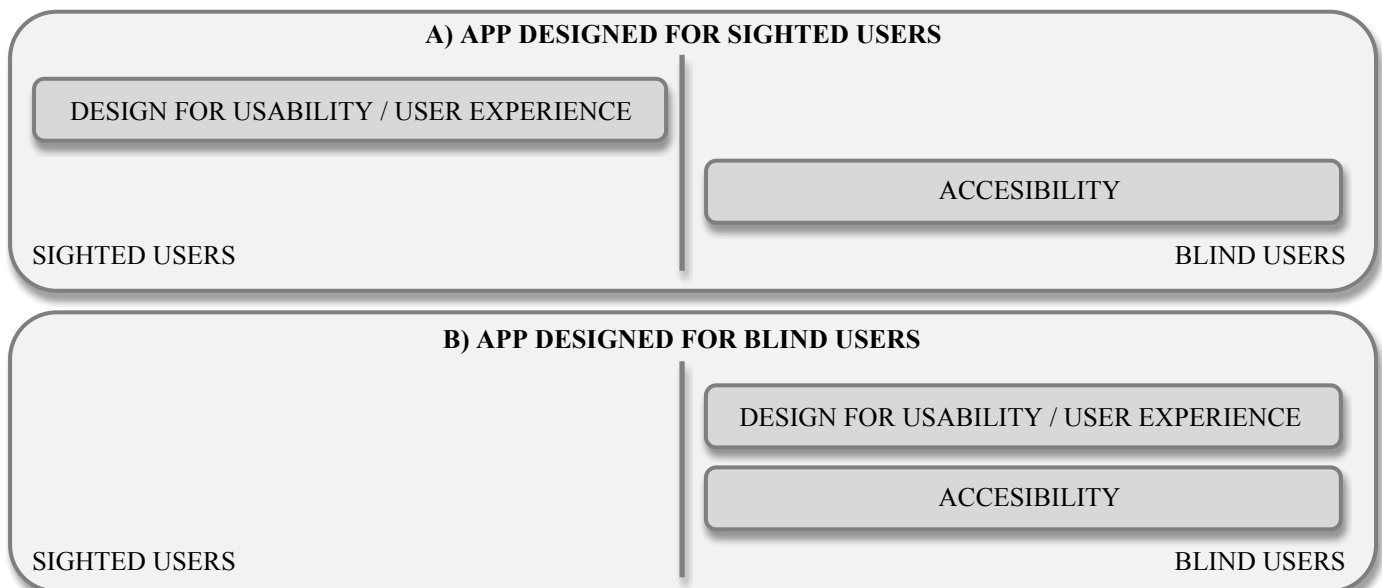


Figure 2. Mobile Apps designed for (a) sighted user (b) blind users.

It is also clear the advantage of using this low vision focused design in the development of specific apps, such as magnifiers, screen readers, GPS locators, etc.

An analogy that would help to understand this design singularity occurs between the iPhone and the iPad devices. If you have to design the same app for both devices, the user interface can be dramatically different. There are even specific iOS controls that are available only for iPad. If the size of the device could influence so much the final appearance of user interface, you can imagine how different would be the App user interface whether the final user is sighted or blind.



Figure 3. Non Usable Accessibility design.

In other words, the scenario given on Figure 2b would be defined as *Usable Accessibility Design*. Design for usability is focused on low vision users and it is using the available accessibility features.

Another useful analogy to understand our approach is shown in Figure 3. It shows a ramp that is used to cross some steps. The initial design of the steps was not conceived for disable people, but the ramp is an alternative that would work. The slope is quite pronounced and the disabled users would require the assistance of other people to cross the steps. The best alternative would have been the design of an alternative accessible route with a lower slope.

If we translate this to the design of mobile apps for visually impaired users, the figure of the ramp would be equivalent to the *Voice Over* feature included on the iPhone and iPad devices. This feature acts like an extra layer added to the mobile app, which was originally designed for sighted users. Voice Over is a great tool that facilitates low vision users the interaction with apps designed for sighted users, but a design centered on visually impaired users would be a much better solution.

### III. LOW VISION MOBILE APP PORTAL

The main goal of our project is to create a collection of mobile Apps for visually impaired users. In order to facilitate the access to these apps, our approach consists on the creation of a *low vision mobile portal*. The appearance of the portal is a simple App that connects the low vision user with the most common apps, such as phone, contacts, messages, notes, etc.

The user can customize the portal, adding or removing apps directly from the Apple Store. Other features such as icons size, colors, screen contrast or voice speed would also be easily customized by the low vision user.

In order to associate an App to the low vision portal, it should comply with a minimum set of specifications. The main requirement is that the App should be specifically designed for low vision users. An App designed for sighted users should be redesigned to meet this requirement and include it in the low vision portal.

Nowadays, there are thousands of apps available on the Apple Store. Although the Apps are divided in different categories, it is not easy for a low vision user to explore all the apps and find an adequate one. The main idea of this portal is to identify those apps and to facilitate low vision user to locate and use them.



Figure 4. Low Vision Mobile App Portal.

.Making mobile apps oriented to blind and visually impaired people is not a novel idea. There are already some mobile apps on the market that include a collection of basic Apps such as phone, contacts, etc. The main difference between these mobile apps and our proposed *Low Vision Mobile App Portal* is that our portal includes own apps and third party apps. Thanks to this portal, low vision users can access easily to a wide variety of Apps specified design for them. On the other side, developers can use this portal to promote their apps between the blind and visually impaired community.

Figure 4 shows an example of the Low Vision Mobile App Portal. It would be defined as a mobile desktop with icons to access apps specifically designed for low vision users. Some of the apps shown in the Figure 4 are Phone, Contacts, SMS, Alarm, Calendar, Battery, Email, Magnifier and GPS Location between others. There is also a setting option to adjust some other features: screen contrast, background, icons size, user interactions, text to speech parameters, etc.

#### IV. MOBILE APPS FOR VISUALLY IMPAIRED USERS

In order to ensure a good user experience, it will be important that all the apps use the same kind of controls to interact with the low vision user. The way the user navigates through views within the app should be similar across different apps.

To make this possible, the first step is to identify those controls used by sighted users that would be also valid for low vision users. A *button* type control seems to be valid for both blind and sighted users. But other controls such as toolbars, segmented controls, tables or data pickets would need to be adapted in order to meet the low vision user requirements. At this stage, new controls for visually impaired users could be also created.

Once we have defined those controls, the next step is to create usability tests for those controls. The tests are used to study how the low vision user interacts with a specific control. On the usability test we can also measure the ability of the users navigating across different views, identifying the controls and interacting with them.

The definition and the test of these controls will give the low vision users the needed user experience that will help them to identify view components and to navigate easily across any app of the portal. Voice Over or any other TTS (text to speech) library would be used as complement to the *usable accessibility design*.

Once we have all the needed ingredients to build a solid low vision user app, the final step is to start with the design of the portal apps. Following are some of the mobile apps that will be included in the basic *low vision mobile portal*. The majority of these apps are traditional apps that have been redesigned to meet the needs of visually impaired and blind users.

- Phone: This App will facilitate the user making phone calls. Figure 5 shows a snapshot of the App. Simple design with buttons and high contrast. As the user moves around the screen touching numbers, the text to speech feature reads those numbers.



Figure 5. Phone App for low vision users

- Contacts: This App will allow the user to navigate through his contact list. Voice over combined with special gestures will make it easier to locate a specific contact in the list.
- SMS: By means of this app, the user will send SMS messages to his contacts in a very simple way.
- Alarm: Date, Time and alarms can be set using this App.
- Calendar: The low vision user will use a special context, which will help him adding and editing notes to the calendar.
- Battery: Text to speech is used to notify the user about the battery level.
- Email: Special email client designed for low vision users.
- Magnifier: The iPhone camera works as a magnifier, facilitating the low vision users the reading of books, newspaper, etc. Special image filters are used to create a high contrast inverting the image to obtain a dark background.
- GPS Location: This simple App will inform the user about his location: street number, city.

- Social network clients: A collection of apps will be developed to allow visually impaired and blind users the connection with social networks, such as twitter or facebook.

In order to help developers to add their own apps to the mobile portal, a specified API will be designed. This API will include the low vision mobile app portal features, which will facilitate developers the design of Apps for low vision users. Design for usability is an essential premise in order to achieve the best user experience.

## V. PRELIMINARY RESULTS

The first stage of the project consisted on the design and development of a low vision library for iOS devices (iPhone & iPad). This library includes several controls specifically designed for low vision and visually impaired users. Two are described:

- *Low Vision Wheel Control*: This control was conceived to allow low vision users to navigate through a collection of data. An example would be the navigation through the telephone contact list. iPhone and iPad devices offer accessibility options to interact with tables, but they require users to use many gestures to get to the desired location. If the users are manipulating a long list it is really annoying to move up and down the list. Our alternative uses a wheel control. The metaphor behind this control is the same that using the wheel button of a traditional mouse. When the user touches the control on the screen, a speech message tells him that it is a wheel control that is connected to a given list. Then, the user touches the screen with a second finger and starts moving it vertically up and down. As the user moves through the list, the system starts speaking the list content. The sensibility of the wheel can be changed. The control is smart enough to detect slow and fast movements of the second finger, which are related to slow or fast movements through the list. This control can be defined vertically or horizontally.
- *Low Vision Segmented Control*: This control was conceived to facilitate low vision users the interaction with a segmented control. This kind of control is used when the user has to select an option from a small list of options. A good example of segmented control would be a list of colors where the user has to choose one color from a list of four colors. The iPhone offers a native “segmented control” that includes accessibility options, but low vision users find it difficult to interact with it. We have designed an alternative that uses a single button to choose a value from a small list. When the user touches the button, a speech message tells him that he has found a segmented control, with a specific value (ex: color “red” selected). To select other values from the list, the user just need to tap with a second finger on the screen until the desired value appears.

Other low vision controls have been designed and tested by a group of low vision users, that have been involved in the design of these controls.

An iOS Low Vision control library will be available to allow developers to use these controls in their apps.

## VI. CONCLUSIONS AND FUTURE WORK

This paper deals about mobile apps design, focusing on design for usability centered on visually impaired and blind people.

The majority of the existing mobile Apps have been initially designed for sighted people. Some of them have been adapted, including accessibility features that allow low vision people to use them. In terms of design, accessibility features are added as a new layer to the apps, acting as a patch that is placed over the initial design layer. iPhone and iPad devices use Voice Over feature to facilitate the accessibility.

Although Voice Over is a power tool to make the apps accessible, it is not enough. Our way of thinking is that the best way to ensure that the app will be usable by low vision users is to design the App specifically for them, instead of designing the app for sighted people and adding extra accessibility features. In order to achieve this goal, special controls (buttons, sliders, tables...) have been redesigned to comply with low vision users needs.

According to this, our research work has lead us to the creation of a collection of Apps, following a common redesign centered on visually impaired and blind users. We have also created a *low vision mobile portal* that includes these apps, facilitating the visually impaired and blind users the access to a wide collection of Apps specifically designed for them.

The portal is open to any developer willing to design an App for low visual users. An API including special low vision controls will be available to facilitate the design of the App in order to achieve the best user experience.

An iOs Low Vision Library is been developed to allow other users to include these controls in their apps and offer them in the low vision portal.

Future work will consist on the use of technology and mobile devices in order to create new products to assist visually impaired and blind users, making their life more accessible.

## ACKNOWLEDGMENT

This research work is supported by IMPIVA (Instituto de Pequeña y Mediana Industria de la Generalitat Valenciana, Spain) [18] as a project in Technology Innovation and Entrepreneurship. The work has been supervised by Raylight Soluciones Tecnológicas [19], a Spanish company that offers products for low vision users.

We want to thank Rosario Vidal from Jaime I University [20] and AIDO (Instituto Tecnológico de Óptica)[21] for their previous research work on this project.

## REFERENCES

- [1] R.L. Mace. “Universal Design”, <http://design.ncsu.edu/alumni-friends/alumni-profiles/ronald-mace>

- [2] M.F.Story, "Maximizing Usability: The Principles of Universal Design", *Assistive Technology* 10:1, 1998, pp. 4-12
- [3] H.Takagi, C. Asakawa, K. Fukuda, J.Maeda "Accessibility Designer: Visualizing accessibility for the blind", ASSETS'04, Atlanta, Georgia, USA, Oct 18-20, 2004
- [4] J. Mankoff, H. Fait, T. Tran, "Is Your Web Page Accessible? A Comparison Study of Methods for Assessing Web Page Accessibility for the Blind", CHI 2005, Portland, Oregon, USA, April 2-7, 2005.
- [5] S. Leuthold, J.A.Bargas-Avila, K. Opwis, "Beyond web content accessibility guidelines: Desing of enhanced text user interfaces for blind internet users", *Int. J. Human-Computer Studies* vol. 66, 2008, pp.257-270.
- [6] K.Fukuda, S. Saito, H. Takagi, C. Asakawa, "Proposing New Metrics to Wvaluate Web Usability for the Blind", CHI 2005, Portland, Oregon, USA, April 2-7, 2005.
- [7] J.P. Bigham, A.C. Cavender, J.T. Brudvik, J.O. Wobbrock, R.E. Ladner, "WebinSitu: A Comparative Analysis of Blind and Sighted Browsing Behavior", ASSETS'07, Tempe, Arizona, USA, Ocy 15-17, 2007
- [8] J. Kjeldskov, J. Stage, "New Techniques for usability evaluation of mible systems" on *Int. J.Human-Computer Studies*, vol 60, 2004, pp.599-620
- [9] A. Kaikkonen, T.Kallio, "Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing", *Journal of Usability Studies*, Issue 1, vol 1, Nov 2005, pp 4-16.
- [10] S. Vidal, G. Lefebvre, "Gesture Based Interaction for Visually-Impaired People", proceedings NordiCHI 2010, Oct 16-20.
- [11] S.K.Kane, J.P. Bigham, J.O.Wobbrock, "Fully Accessible Touch Screens for the Blind and Visually Impaired", University of Washington.
- [12] M. Hassenzahl, N.Tractinsky, "User experience – a research agenda", *Behavior & Information Technology*, Vol. 25, No.2, March-April 2006, pp 91-97
- [13] B.Buxton, "Sketching user Expeciences: getting the design right and the right design", Morgan Kaufmann Publishers, Elsevier, 1997
- [14] "User Experience definition" on wikipedia , [http://en.wikipedia.org/wiki/User\\_experience](http://en.wikipedia.org/wiki/User_experience)
- [15] "iOS Human Interface Guidelines. User Experience", Apple Inc. 2011
- [16] "Accessibility definition" on wikipedia , <http://en.wikipedia.org/wiki/Accessibility>
- [17] "Accessibility Programing Guide for iOS. User Experience", Apple Inc. 2011
- [18] <http://www.impiva.es>
- [19] <http://www.raylight.es>
- [20] <http://www.uji.es>
- [21] <http://www.aido.es>



## Building Bridges Between Elderly and TV Application Developers

José Coelho, Carlos Duarte, Pedro Feiteira, Daniel Costa and David Costa

LaSIGE, Department of Informatics

University of Lisbon

Lisbon, PT

{jcoelho, cad, pfeitoira}@di.fc.ul.pt, {thewisher, dcosta}@lasige.di.fc.ul.pt

**Abstract-** The development of new digital TV systems and the design practices adopted in the development of new TV based applications often isolate elderly and disabled users. By considering them as users with special needs and not taking their problems into account during the design phase of an application, developers are creating new accessibility problems or just keeping bad old habits. In this paper, we describe a novel adaptive accessibility approach on how to develop accessible TV applications, by making use of multimodal interaction techniques and without requiring too much effort from the developers. By putting user-centered design techniques in practice, and supporting the use of multimodal interfaces with several input and output devices, we confront users, developers and manufacturers with new interaction and design paradigms. From their evaluation, new techniques are created capable of helping in the development of accessible TV applications.

**Keywords**—multimodal; adaptation; developers; elderly.

### I. INTRODUCTION

Ageing is certainly an obstacle to adequate human-computer interaction, mostly because of physical and cognitive impairments. Traditional computational systems only provide keyboard and mouse interaction to users. This makes impossible, for example, for users with severe motor impairments to interact in any manner (at least effectively). Also, as recent developments are responsible for new television (TV) systems and applications, unimodal interaction is still being favored without accessibility concerns, excluding persons whom suffer from an impairment of the sensory channel needed to interact. This situation brings social exclusion and e-exclusion to the Human-Computer Interaction (HCI) world and new TV platforms, as it seriously restricts actions and information access to users with impairments (like the elderly), providing means of interaction exclusively for the so called “normal users”. However, multimodality can resolve this issue by offering the possibility of presenting content in many ways (audio, visual, haptic), and in the most suitable way to each user’s characteristics. Also, by offering users the possibility to use the inputs more adequate to them (or the context of interaction), in a single or combined manner, multimodal interaction can improve interaction efficiency and, more importantly, accessibility.

Multimodal interfaces aim to provide a more natural and transparent way of interaction with users. They have been able to enhance human-computer interaction (HCI) in many numbers of ways, including: User satisfaction: studies revealed that people favor multiple-action modalities for

virtual object manipulation tasks [14]; Oviatt [17] has also shown that about 95% of users prefer multimodal interaction over unimodal interaction; Robustness and Accuracy: “using a number of modes can increase the vocabulary of symbols available to the user” leading to an increased accessibility [15]. Oviatt stated 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” [18]; 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 [16]; Adaptivity: Multimodal interfaces also 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 [15]. However, Vitense [20] illustrates the need of additional research in multimodal interaction, especially involving elderly people. This paper tries to extend this knowledge.

Also, the majority of current approaches to the development of multimodal or adaptive systems, either addresses specific technical problems, or is dedicated to specific modalities. The technical problems dealt with include multimodal fusion [10], presentation planning [10], content selection [12], multimodal disambiguation [18], dialogue structures [3], or input management [9]. Platforms that combine specific modalities are in most cases dedicated to speech and gesture [19], speech and face recognition [11] or vision and haptics [13]. Even though the work done in tackling technical problems is of fundamental importance to the development of adaptive and multimodal interfaces, it is of a very particular nature, and not suited for a more general interface description. Also, frameworks supporting the development of interfaces for various devices exist; however, they do not consider the specificities of multimodal interaction in its design [5][6]; or they focus only on the use of the same modality in different devices [1]; or they ignore the possibility of adapting the components properties and features in run-time placing the burden on the designer [4]. In general, they do not consider in there architectures the introduction of modalities, and how they can be explored to achieve the goals of Universal Access.

In the following, we first explain how European funded project GUIDE [7], aims to adapt interaction and UI presentation to fit each user’s characteristics and level of expertise. Also, resulting from specific user trials and discussions with developers, we also show how it makes use

of a User Initialization Application to know and instruct its users, and how it supports adaptation by providing developers with solutions for UI modification, and tools for helping in the development of new user-centered and accessible applications. All this attending to user needs and differences, at the same time as it takes into consideration the developer's interests.

## II. CHARACTERISTICS OF GUIDE PROJECT

### A. End-Users and Goals

GUIDE [7] aims to achieve the necessary balance between developing multimodal adaptive applications for elderly and disabled users, and preserving TV and Set-Top Box(STB) developers/manufacturers design methodologies and efforts. Consequently, there are clearly two different end-users of this project: elderly and impaired users and developers of TV based applications. Creating a bridge between these two, we have also the STB manufacturers who dictate the rules about which type and which characteristics of applications can be used on a TV based environment. Firstly, for elderly and users with impairments, GUIDE has the goal of providing new ways of interacting with a TV, by applying multimodal interaction, supporting the use of different devices as well as different combinations of input and output techniques, and adaptation to each application's UI an each user's way of interaction. In other words, elderly or impaired users who are having difficulties interacting with modern TV systems because of their complexity, will be able to interact in a more intuitive way, using alternative modalities in a single or combined fashion, while each interface characteristics will also be adapted to fit user's characteristics automatically. For all this, GUIDE has as a clear defined target environment, a STB connected to a TV in user's home (and closed) environment. Secondly because developers tend to have no concerns about accessibility when designing TV applications, GUIDE has to be capable of reducing development effort in a radical manner. For that end, GUIDE wants to create a toolbox for accessible applications and UI design, shifting the design principle from a conventional user-centered design process to a GUIDE-assisted design and development process. Through all this, GUIDE also wants to ensure that developers (and also manufacturers) can maintain the control over the modifications made on their own applications UI. Meaning, the adaptation provided by the system for adapting interfaces to user characteristics must have boundaries that cannot be crossed. And these boundaries are defined by the developers.

### B. Multimodal Interfaces and Devices

Input modalities to be supported in GUIDE are based in the more natural ways of communication for humans: speech and pointing (and gestures). Complementary to these modalities, and given the TV based environment, the framework should support the usage of remote controls and other devices capable of providing haptic input or feedback. As a result, 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 should consider and integrate the following output components: 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. A tablet may also be used to clone the TV screen or complement information displayed on the TV screen but essentially is used as a secondary display. The main user interface should be able to generate various configurable visual elements such as text (e.g., subtitles or information data), buttons for navigation purpose, images and video (e.g., video conference or media content). Additionally also a 3D avatar is generated and 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.

In order for the UI to be adapted to the user's needs, these elements are necessarily highly configurable and scalable (vector-based). Size, font, location, and color are some attributes needed to maintain adaptability. These graphical elements enable the system communication with the users by illustrating, answering, suggesting, advising, helping or supporting them through their navigation. Also, both input and output modalities can be used in a combined manner to enrich interaction and reach every type of user.

### C. Discussion: What GUIDE needs to know

For reaching its goals, GUIDE has to define a framework structure and collect information by asking and testing its end-users. So, the following questions have to be answered: What components the GUIDE framework has to have? What are the main preferences and typical behavior of elderly users when interacting with the system, and how to collect these preferences? How to perform automatic UI adaptation? How to help developers and manufactures in design process?

## III. LEARNING FROM END USERS

To get answers to the questions above, we firstly derive end user requirements from results obtained through a quantitative and qualitative analysis of data recorded in comprehensive user trials [8]. Secondly, we organized focus group sessions with developers and used an online survey as qualitative research tools in gathering additional requirements from developers and STB platform providers

### A. Initial User Trials

The GUIDE project pursues a User Centered Design (UCD) process, taking into account that one of the main principles that characterize UCD is iterative design. According to this principle the system is designed, modified and repeatedly tested. This iterative cycle allows the designers to think in the product design and include the changes needed depending on the users' feedback. Following this approach, an initial study to elicit user requirements has been carried out.

1) *Main Objectives*

Additionally to the identification of viable usage methods (gestures, command languages) of novel traditional UI paradigms for the different impairments in the target groups via user studies in realistic user scenarios, this user trials also have the goal to generate quantitative and qualitative user data in order to establish and construct a generic user model. This user model will provide data representations for each user and will constitute the first step for adaptation in GUIDE, and will “virtualize” user impairments to try to capture the amount of knowledge needed for application design.

2) *Organization and Setting*

The initial user studies carried out can be divided in two different categories; one survey session and one technical trials session. While the aim of the survey was to collect qualitative information about application acceptance, user habits and modalities of interaction, the objective of the technical trials was to gather both quantitative and qualitative data and observe the interaction between the elderly and the system, performing simple tasks in the context of TV interaction. In [8] you can find an extensive description of the tests performed.

B. *Developer Focus Groups and Survey*

The GUIDE system is not exclusively focused on elderly users, but also centered in developers of TV based applications and manufacturers of STBs. For this reason, major discussions regarding subjects like adaptation, elderly user’s interaction, type of applications, and developers requirements for making possible the GUIDE ideas, has taken place in this evaluation, by performing both focus group with these end-users and by launching an online survey with the same user target.

1) *Main Objectives*

The general goal is to explore and understand the common practice among developers working on STBs. Thus the first objective is to gain data about current tools and APIs used in Set top box/connected TV platforms and to investigate how accessibility is currently perceived and applied in the industry. Secondly, exploring developer knowledge to identify which tools would developers need to efficiently integrate GUIDE-enabled accessibility features into their applications. Additionally, stimulate new ideas through discussions and to identify new relationships between objects embodying GUIDE concepts and objects embodying common practice. And finally, inform STB application development community about GUIDE.

2) *Organization and setting*

Developer Focus Groups: Two focus group sessions were carried out with connected TV platform providers and developers of applications and user interfaces deployed on STBs in a natural and interactive focus group setting. The sessions were conducted by two moderators (for ensuring progress and topic coverage) and each focus group session had between six and eight participants and lasted between 120 and 150 minutes. Sessions were initiated with presentations of scripts containing development and use cases that cover different aspects of the GUIDE project and

its concepts. Presentations of each development case script lasted 10 minutes and were followed by 30 minutes of interactive brainstorming, and discussions.

Developer Online Survey: A questionnaire was designed to investigate how accessibility is currently perceived and applied in the industry. In addition, the survey was used as a medium to let respondents vote on the most important features of the envisaged GUIDE framework and toolbox.

Both survey and focus group were composed by the following participant types: STB test developers, STB experts in Innovative part, Flash application developers, HTML developers, middleware STB developers, architects in STB platforms, GUI developers for STB, project managers for STB projects, managers in Innovative projects for STB, product and marketing managers, research community, and standardization bodies and related organizations. In total, 81 participants from 16 countries, and 30 companies all over the world, participated.

C. *Results and Conclusions*

From the realization of both initial user-trials and developers focus group (and online survey), we now summarize qualitative results which will work as starting points for the next section of this paper:

1) *User Survey Results*

The large numbers of variables contained in the data set were submitted to a two-stage process of analysis where correlations were made and a k-mean cluster analysis [2] was performed, reducing the results to only significant data. Resulting from this, 3 user profiles capable of discriminating differences between users were created. These profiles were formed by combining and grouping all modalities simultaneously such that a specific grouping may represent capability on users perceptual, cognitive and motor capability ranges. The main differences noticed were the following measures: capability to read perfectly from close and distant vision; capability of seeing at night, and color perception; capability to hear sounds of different frequencies and to distinguish conversations in a noisy background; cognitive impairments; and mobility diagnosis like muscular weakness and tremors.

2) *Technical User Trials Results*

Big, centered and well-spaced buttons were preferred by users because they are easier to see and select (and elderly users typically have some kind of visual and motor impairments). Additionally, users prefer medium sized fonts and medium volumes for audio, but users with impairments tend to prefer bigger fonts and higher volumes. However, more than based on user abilities or preferences, both visual and audio elements configuration, depends on the interaction context and must be at all times modifiable and repeatable by the user. All the preferences described regarding visual components, reflect the low efficiency (lot of time needed for each selection) and accuracy (wrong target when selecting) registered when interacting with any type of pointing in these tests.

Users clearly preferred gestures easier to make (swipe and pinch), and have no problem whatsoever interacting by gestures. It was also evident that alternative ways of

interacting with the TV (speech and finger pointing) are preferred to the traditional way. Also, training makes any type of modality more efficient as the user learns what is required to perform each interaction. However, any type of interaction should not be imposed on the users, but be available as an intuitive option for interacting with the TV. Additionally, when not used by the user intuition, modalities of interaction should be explained to the user before he or she starts using the system.

Every user is able to interact multimodally with the system and combine speech and pointing, even when they prefer only one modality. Users exhibited different multimodal interaction patterns during the trials and there is no specific interaction pattern for each user (a user can speak first and point afterwards, and in the next interaction do the opposite). Users can also change the way they interact depending on the type of feedback given while interacting. Regarding user preferences in input and output modalities, there are clear differences between what users say they prefer, and what users really ask for when interacting. In fact, 100% of the users want multimodal output every time information is presented to them, because every user who said to prefer only one type of feedback admitted differently when in specific interaction contexts. The same happened concerning input modalities, with almost half of the users admitting, when confronted with practical tasks, that they were wrong when they said to prefer only one modality.

The results obtained in these trials enforce the need of a multimodal system and also the need for adaptation, as we can see in a more detailed fashion in [8].

### 3) *Developer Focus Groups and Survey Results*

Developers agree that if users are involved in every development phase of the applications (or in the maximum phases possible), the resulting UI will be more usable. It was concluded that for elderly people UIs should be maintained clear and simple, however without giving the impression that it has been designed for someone with impairments (not leaving the feeling of a “system for seniors”). Additionally, costs are the current major reason for reduced application of user-centered design in the industry (followed by time and lack of awareness). As the current most important device on interaction with STBs, the remote control must continue to have a central role in the interaction, and should only be relegated to a secondary role if that is a result of each user interaction preferences. Gesture control and speech input are recognized as secondary technologies. In general, participants agree that automatic adaptation of user interfaces can help elderly users to access ICT services. However GUIDE adaptation mechanism should never change interface aspects unless it is mandatory for specific user interaction. Also, radical changes in the UI must be avoided so that the user feels he/she is in control and not get lost in the interface. If a radical change is indispensable the UI must inform the user of the proposed changes. Identified as the main obstacle to UI adaptation is the fact that elderly users present too many differences between each other. Therefore, for adaptation to fit each user, GUIDE has to first find a way to know his or her impairments, preferences or characteristics. This “discovery” will have to occur the first time the user interacts

with the system, and will have to be short, not too much intrusive and entertaining to the user. The most important conclusion debated in this subject is the one saying GUIDE should support UI mark-up as interface between application and GUIDE adaptation. This way, developers will be allowed to keep tools and development environments and without too much additional effort, take a first step to accessible design. Web developers mostly use HTML editors as the most important tools in Web & TV development. However, having to learn new development processes will drive developers away from the GUIDE framework. So, developers should not be required to develop taken into consideration specificities of the multimodal operations but have a clear specification of how such devices interact with the framework. As it was already described in UI adaptation results, identification of UI components should be made using only mark-up language, however applications coded using dynamic HTML (through JavaScript) must continue to be able to change, remove or insert elements in the currently rendering page. Meaning, all changes in application presentation will need to be identified at run-time. For most participants connected TV platforms and STBs will be most relevant platforms in the future. Also, Web-based application environments will become more important for Web & TV. Manufactures stated increasing STBs capabilities cannot raise its price to much, or development will be more difficult and costly. Developers also pointed out GUIDE system must consider situations where multiple users are using the TV and services.

## IV. MULTIMODAL APPLICATION DEVELOPMENT

From the results and implications reported in the previous section of this paper, we now derive GUIDE project solutions for giving answers to the same questions raised in the beginning of this paper.

### A. *Multimodal and Adaptive Framework*

We now give an overview of the GUIDE framework [7] following an interaction cycle, starting from the user input and going through the construction of the system’s output to be presented to the user.

A user provides input through multiple devices and modalities which can be used simultaneously. The signals from recognition based modalities are processed by interpreter modules (e.g., a series of points from the motion sensor go through a gesture recognition engine in order to detect gestures). The signals from pointing modalities go through input adaptation modules (e.g., in order to smooth tremors from the user’s hand). Both interpreter and adaptation modules base their decisions on knowledge stored in the user profile, thus improving the efficiency of noise reduction in the input signals. Then, the multimodal fusion module receives, analyses and combines these multiple streams (outputs of input interpreters and input adaptation modules, or raw data that did not go through any of these) into a single interpretation of the user command based on the user, context and application models. This interpretation is sent to the dialogue manager who decides which will be the application’s response, basing its decision on knowledge

about the current application state and the possible actions that can be performed on the application in that state. The dialogue manager decision is fed to the multimodal fission module, which is responsible for rendering a presentation in accordance to which output to present (derived from the application itself and the application model), the user abilities (accessed through the user model) and the interaction context (made available through the context model). The fission module takes all this information and prepares the content to render, selects the appropriate output channels and handles the synchronization, both in time and space, between channels when rendering. This rendering is then perceived by the user, which reacts to it, and starts a new cycle by providing some new input.

### B. User Initialization Application

In both technical user-trials and focus groups, it is the necessity of knowing every user characteristics, preferences and impairments from the first time he or she interacts with the system. This is mandatory because of the user's differences and the necessity of adapting both UI components and interaction to fit each user, as well as the necessity of instructing the user about every possibility of interaction in order to reach the maximum efficiency when using the system. GUIDE adaptation begins through a User Initialization Application (UIA) that allows for the acquisition of primary assumptions about the user. So, knowing that each user model contain assumptions about interesting characteristics of user subgroups, after "going trough" the UIA, a user is assigned to a user model as certain preconditions are met. From that moment on, and for any GUIDE application the user interacts with, the system is "initially" adapted to him/her. It's relevant to say that the UIA is presented to the user as a simple step-by-step configuration of a "general" interface. In each step, different types of contents and different contexts of interaction are presented, so the user can test different components and parameters, and the system learns the user characteristics, from his impairments to his preferences. Addressing the results from the developer focus groups, every UIA run as to be short in time, intuitive and transparent to the user and also serve as a "tutorial" for learning every modality of interaction available in the system. Additionally, the user must be recognized (facial or voice patterns) by the system so that the information provided can be stored in a profile and loaded every time the user interacts with the system.

### C. Simulation of User Impairments

As developers need tools for helping saving time and cost in the development of inclusive TV base applications, GUIDE offers a simulator [2] which will allow the developer to perform accessibility tests based on virtual users, saving much time in comparison to tests with real users. So, evaluation as a typically expensive step in user centered design is supported in GUIDE by a simulation functionality allowing to illustrate to developers how users with typical impairment profiles will perceive or may interact with an application. The simulator can show how certain visual and strength impairments influence the way a user perceives and

visualizes a certain UI (e.g., how an elderly color blind user sees a specific UI), and also what are the effects of those impairments in the user interaction (e.g., predicting cursor paths on the screen or task completion times). This simulator can be characterized as a tool for helping developers to take adaptation into consideration at design time.

### D. Filtering

As verified by the inefficient and erroneous use of pointing interaction when performing selections in the user trials, elderly users potentially have a wide range of impairments that hinder their ability to communicate their intentions to an application. In some cases these impairments can be severe, and significantly affect the speed and accuracy. This leads to an inefficient or even undesirable interaction with an application. The use of cursor smoothing techniques in GUIDE consists in processing the raw user input to obtain a filtered input (Input Adaptation Module described in the framework). This requires the usage of efficient statistical signal processing schemes to estimate the user's intended operations in real time. Basically it consists in the application of corrective forces and forcing relatively smooth paths in a cursor interaction as well as assigning attraction fields to UI elements. Therefore, the following graphical UI filters can help improving pointing interaction within the GUIDE project:

Exponential averaging: this modification calculates the cursor position  $p_i$  as  $p_i = \alpha x_i + (1-\alpha)p_{i-1}$ , where  $x_i$  is the user input,  $p_{i-1}$  is the previous cursor position and  $\alpha \in [0,1]$  is a parameter determining how strong the user input influences the cursor position. This method produces smooth cursor traces but has the drawback that it can produce a delay between user's intended position and the actual position;

Damping: This method introduces a quadratic force that opposes the velocity of the cursor preventing sudden changes in directory or speed when interacting; Gravity well: This method warps the cursor space, generating attractive basins to ease the selection of visual targets. This simplifies pointing interaction selection forcing the selection of buttons or UI elements that are more close to the location where the user is pointing.

Considering the different user characteristics and impairments, and the different UI element configuration, the existence of these filters make possible that motor impaired users can more easily interact with pointing and also makes possible the use of small and less spaced buttons in applications UIs avoiding errors in selection caused by the proximity of the buttons. All because pointing interaction accuracy is raised.

### E. Semantic Programming and Run-Time Adaptation:

The specification of TV based applications in GUIDE will be based on Web-based languages like HTML, CSS and JavaScript because of their wide acceptance among developers and compliance with STB specifications. However, in GUIDE exists the additional side-condition of specifying multimodal applications that needs to be merged with these web-based specification languages. This is made by specifying additional information about how an

application is supposed to adapt in different modalities. For this semantic annotations are added to the HTML code, based on the WAI-ARIA draft specification of the W3C. Only by providing this type of supplementary information it is possible for the system to create an abstract representation of the application. Then, using an automatic application transformation module the system converts the annotated application description into a modality-independent application representation, the Application Model described in the framework. Subsequently, and depending on the user interacting and on the level of control defined by the application developer, adaptation of UI components is performed. Developers can create their applications and UIs in an established manner, and GUIDE automatically adapts the UI to the user. This avoids having to design many user interface templates for various heterogeneous user groups. Therefore, GUIDE provides the application developers with three possible levels of adaptive control: **Augmentation**: presentation and interaction options taken by the developer are not subject of change. Instead, if the user model suggests that the presentation is insufficient for the user abilities, the presentation is augmented in different modalities (for example supplementing a visual interface with sound feedback). The multimodal fission mechanism renders the application output directly, augmenting or not the rendered presentation depending on the user model; **Adjustment**: application rendering is adjusted to the abilities of the user (for example adjusting components of a visual interface to fit user characteristics, like raising font size or button size). The rendering changes can be achieved through CSS manipulation. Adjustment can be combined with augmentation; **Replacement**: application content is replaced by a new version of content. In this level the application hands over completely the rendering responsibilities to the multimodal fission component, and the UI is rendered by transformation of application content, layout configuration, as well as presentation styles. Augmentation and adjustment can also be performed.

## V. CONCLUSIONS

The results obtained in the technical user-trials about the existence of disparity between what modalities users say they need, and what modalities they ask for when using the system, favors multimodality almost every time. This only helps to prove that the use of several input and output modalities is indispensable in the development of multimodal TV based applications for all. Also indispensable, are the components identified in the GUIDE framework, and the combined use of semantic programming and run-time adaptation mechanisms to fit UI components to each user characteristics. Additionally, the use of a simulator of user impairments can help developers understand at design-time how certain UI templates and components are perceived by different users with different impairments, preventing user exclusion and making accessible applications easier to design

## ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 248893.

## REFERENCES

- [1] Balme, L., Demeure, A., Barralon, and N., Coutaz, J. & Calvary, G. CAMELEON-RT: A Software Architecture Reference Model for Distributed, Migratable, and Plastic User Interfaces. In EUSAI'2004
- [2] Biswas, P., Robinson, P., and Langdon, P.: Designing inclusive interfaces through user modelling and simulation. *International Journal of Human Computer Interaction*.
- [3] Blechschmitt, E., and Strodecke, C.: An architecture to provide adaptive, synchronized and multimodal human computer interaction. In MULTIMEDIA '02, NY, USA, pp. 287-290.
- [4] Bouchet, J., and Nigay, L.: ICARE: a component-based approach for the design and development of multimodal interfaces. In CHI '04, NY, USA, pp. 1325-1328.
- [5] Calvary, G., Coutaz, J., and Thevenin, D.: A unifying reference framework for the development of plastic user interfaces. In EHCI '01, London, UK, pp. 173-192.
- [6] Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Souchon, N., Bouillon, L., Florins, M., and Vanderdonck, J. (2002). Plasticity of user interfaces: A revised reference framework. In TAMODIA '02, Bucharest, pp. 127-134.
- [7] Coelho, J., and Duarte, C.: The Contribution of Multimodal Adaptation Techniques to the GUIDE Interface. HCII2011, Orlando, Florida, USA, pp. 337-346.
- [8] Coelho, J., and Duarte, C., Biswas, P., Langdon, P. Developing Accessible TV Applications, Proceedings of ASSETS 2011, pp. 131-138.
- [9] Dragicevic, P., and Fekete, J.D.: The input configurator toolkit: towards high input adaptability in interactive applications. In AVI '04, ACM Press, NY, USA, pp. 244-247.
- [10] Elting, C., Rapp, S., Mohler, G., and Strube, M.: Architecture and implementation of multimodal plug and play. In ICMI '03, ACM Press, NY, USA, pp. 93-100.
- [11] Garg, A., Pavlović, V., and Rehg, J. (2003). Boosted learning in dynamic bayesian networks for multimodal speaker detection. Proceedings of IEEE 91, pp. 1355-1369.
- [12] Gotz, D., and Mayer-Patel, K.: A general framework for multidimensional adaptation. In MULTIMEDIA'04, NY, USA, pp. 612-619.
- [13] Harders, M., and Szekely, G. (2003). Enhancing human-computer interaction in medical segmentation. Proceedings of IEEE, 91, pp. 1430-1442.
- [14] Martin, J.C., Julia, L., and Cheyer, A.: A theoretical framework for multimodal user studies. In CMC98, Tilbur, Netherlands, pp. 104-110.
- [15] Oakley, I., Brewster, S. A., and Gray, P. D.: Solving multi-target haptic problems in menu interaction. CHI'01, Seattle, USA, pp. 357-358.
- [16] Oviatt S. L. Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction*, 1997, pp. 93-129
- [17] Oviatt, S. L., DeAngeli, A., and Kuhn, K. Integration and synchronization of input modes during multimodal human-computer interaction. CHI '97, New York, USA, pp. 415-422.
- [18] Oviatt, S.L.: Mutual Disambiguation of Recognition Errors in a Multimodal Architecture. CHI'99, Pittsburgh, USA, pp. 576-583.
- [19] Sharma, R., Yeasin, M., Krahnstoeber, N., Rauschert, I., G., Brewer, I., Maceachren, A.M., and Sengupta, K. (2003).

Speech-gesture driven multimodal interfaces for crisis management. *Proceedings of IEEE*, 91, pp. 1327-1354

- [20] Vitense, H. S., Jacko, J. A., and Emery, V. K. Multimodal feedback: An assessment of performance and mental workload. *Ergonomics* 46, 2003, pp. 66-87.

## Analysis of Volumetric Tactile Symbols Produced with 3D Printing

Jaume Gual  
 Dept. of Industrial Systems  
 Engineering and Design  
 Universitat Jaume I, UJI  
 Castellón, Spain  
[jgual@esid.uji.es](mailto:jgual@esid.uji.es)

Marina Puyuelo  
 Dept. of Graphic Expression  
 in Architecture  
 Universitat Politècnica de  
 Valencia  
 Valencia, Spain  
[mapuca@ega.upv.es](mailto:mapuca@ega.upv.es)

Joaquim Lloveras  
 Dept. of Engineering Design  
 Universitat Politècnica de  
 Catalunya  
 Barcelona, Spain  
[j.lloveras@upc.edu](mailto:j.lloveras@upc.edu)

**Abstract**— The morphological elements of design used for designing tactile maps and symbols for visually impaired users are points, linear, and areal elements. One of the main characteristics of these elements is their two-dimensional and graphic nature. However, since three-dimensional design came to be a fourth group of elements, volumetric elements, has come into use. The key questions of this study are: Is it possible to extend the range of a discriminatable set of symbols by using volumetric elements with height contrast extended in the Z axis? Can some formal variations of these volumetric symbols be distinguished using the sense of touch? The results of this study show that some tactile symbols with simple volumetric forms are easily recognizable using the sense of touch. In the absence of further studies, this could suggest an affirmative answer to the first research question posed.

**Keywords**—tactile symbols; tactile maps; inclusive design; visual impairment

### I. INTRODUCTION

This study presents the first results of a doctoral thesis carried out on tactile maps and symbols for the guidance of the visually impaired. The work reported here has been included in a research project from the Universitat Politècnica de València (project DPI2008-03981/DPI), Spain).

Tactile maps, as tangible graphic resources, are a group of devices showing graphic information using relief. Tactile symbols are contextualized within this type of device and are normally used with their corresponding keys. These products help blind people understand the features of the environment around them through the sense of touch.

The use of these devices is almost always combined with audio descriptions to facilitate a correct understanding of the tactile exploration. This means that these resources are not used independently. Thus, the challenge is focused on designing more user-friendly and efficient tactile maps. To achieve this goal it is vital to improve two aspects of the information on these devices, on one hand, user-friendliness [1], and on the other, the interaction with sound outputs. This study focuses on improving the former.

All graphic information can be expressed using different elements in relief. Thus, tangible graphics can represent all kinds of graphs, maps, plans, etc, which are difficult to express through text. As will be shown, there are three

elements used to compose the tangible graphics: points, lines, and areas. However, a fourth category of design elements, volumetric elements, are barely used in the design and production of this type of devices, due partly to the conditions of traditional production systems: microencapsulation and thermoforming [2]. The novelty of this work is the study of some possible applications of volumetric elements, basic prisms, specifically in their application as tactile symbols, to improve the usability of tactile maps. The manufacturing technique used in this case to make these symbols is 3D printing which can produce more complex geometries than traditional methods [3].

In contrast, one of the fields of knowledge which has focused most on the issue of tactile devices is that of Cartography and Geography. The integration of other areas, such as Psychology and Education Science, has also paid serious attention to this issue.

From the perspective of product design, it is possible to observe how these products have barely been addressed in depth, despite the fact that the philosophy of Inclusive Design [4] seems an ideal framework to study these objects.

This paper has the following structure:

- Section I. Introduction;
- Section II. Tactile Perception, Typology and Particularities of Relief Maps;
- Section III. Study Description;
- Section IV. Results, and
- Section V. Conclusions and Further Work.

### II. TACTILE PERCEPTION, TYPOLOGY AND PARTICULARITIES OF RELIEF MAPS

On the other hand, tactile perception is a relatively new field of study. David Katz was a pioneer and in 1925 published a classic monograph on the subject, *Der Aufbau der Tastwelt* (The world of touch), which laid the foundation for later studies [5]. These early studies give us a better understanding of the attributes and characteristics of the sense of touch and enable us to design more efficient tangible graphics.

There are different types of tactile maps used for communicating and teaching geography, and also orientation skills, to facilitate movement through certain environments. According to Edman [6], these products can be classified as:

- Mobility Maps;



- Topological Maps;
- Orientation Maps;
- General Reference Maps, and
- Thematic Maps.

It should be noted that the usefulness of these types of devices in facilitating mobility, spatial orientation, and the autonomy of visually impaired people has been clearly demonstrated in previous studies [7][8][9].

However, reading a tactile map also depends on the skills, strategies of exploration, experience, and training of the people using it [10]. These factors allow blind users to recognize the information offered in a tactile product more accurately and effectively, even in real contexts [11]. Another important aspect in blind users is haptic memory, since a blind person explores tactile graphics in a sequential way. In contrast, the phenomenon of visual perception is simultaneous and less time is required to assimilate the same amount of information [12]. This means that the design of tactile devices should be simple, with less information than in the visual version.

Finally, the contexts for the use of tactile maps can be varied, depending on map format and user preferences. These can be previously used at home or at a specific location with the support of a Mobility Instructor. There are even some formats that are portable and can be used in situ. Some studies show that blind users prefer to use them at home, in their own time [13].

#### A. Inclusive Design and Usability

Generally, in order to make this type of product easy to use, and taking into account that tactile perception is not as sharp as visual perception, any tactile-graphic device must contain synthesized information in order to ensure it is easily legible using the sense of touch. If the tactile device includes corresponding visual information, such as colour contrasts or large type, adapted to the specific requirements of other groups, the number of users that may benefit from it may grow to include, for example, the elderly or partially sighted, in keeping with the philosophy of Inclusive Design [4].

There are general requirements that must be mentioned now in order to acknowledge the specific nature of the design process for these products for the sense of touch. However, there are no set criteria and the general requirements greatly depend on the specific experience of each designer.

The maximum size of any tangible graphic must be designed taking into account the space needed using both hands together. A comfortable hand position would include an area approximately the size of an A3 sheet, although maps may be bigger or smaller based on the different formats and types of information to be represented. On the other hand, the scale is conditioned by the constant dimensions of Braille code and the purpose and type of information to be represented. The minimum distance between the elements represented, such as the symbols of a map, must be carefully designed. A minimum separation of 3 mm is needed between elements so they can be recognized using the sense of touch. In any case, these data only represent a small part of all the requirements studied for the design of a correct tactile map.

Extended and more detailed information can be found in reference publications [6] [14] [15].

However, most of the design guidelines published for relief maps are focused on the traditional methods of production, rather than on the techniques used in this study: 3D printing from Rapid Prototyping (RP).

#### B. Manufacturing systems of tactile maps. Rapid Prototyping

There is extensive literature on the manufacture of tactile maps in the field of geography. The usual methods of production are thermoforming of plastic sheets (Fig. 1) and microencapsulation (Fig. 2) [2] [16].

It is important to mention the possibilities opened up by RP for blind people, producing pieces from virtual Computer Aided Design (CAD) models. According to some studies, three-dimensional configurations can improve visually impaired people's understanding of these products [17].

In any case, these techniques enable the production of single pieces or small series in a relatively short time compared to traditional systems of production. Hence their name of 'rapid'. At the same time, they can be used in the early stages of production as master models, that is to say, as preliminary prototypes for long series.

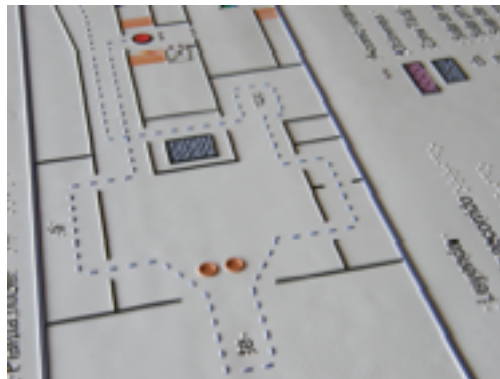


Figure 1. Thermoform copy.

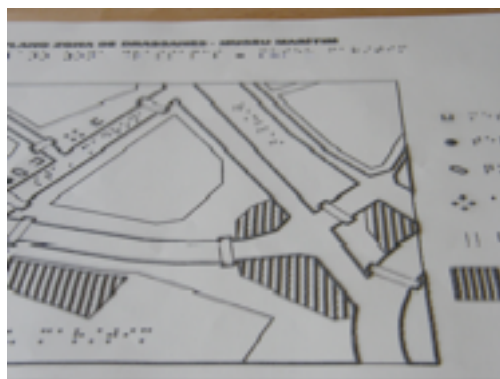


Figure 2. Microencapsulated copy.

But, the main novelty of techniques, such as 3D printing, for the production of tactile maps is the possibility of introducing complex geometries and polychromed pieces. In contrast, the usual techniques employed in the production of

tactile maps, thermoforming and microencapsulation, are somewhat limited in this respect. One of the main disadvantages is their poor surface quality, which makes the model rough to the touch, although this can be corrected with subsequent surface treatments, although so far this implies high costs when producing long series. There have been several previous experiences in the field of haptic devices, especially for the field of tactile scale models [18] [19] [20], but also with tactile maps [21].

### C. Morphological elements for designing tactile maps

According to the literature, there are three morphological elements for designing tactile maps, tactual symbols, or any tangible graphic for the visually impaired [6] [22] [23]:

- Points
- Lines
- Areas

But, if we concentrate on the fundamentals of design, it is apparent that these three types of elements are also called visual or conceptual design elements [24] [25]. Specifically, they are employed in the world of graphic design. The common denominator of these elements is their two-dimensional nature, that is to say, the fact that they are expressed in visual formats within the two dimensions of a flat framework.

However, it is also known that when product designers work in a three-dimensional space, i.e., when objects are represented in relief or in three dimensions, light and volume play a special role. Light models dark bodies and allows a clear perception of volumes in space. Thus, volume is commonly known as the fourth element of conceptual design and can be perceived using the sense of touch.

### D. Symbolology in relief maps

The symbolology of tactile maps has been widely studied in the above mentioned disciplines, particularly Cartography [26]. Recognition, legibility, and discrimination of symbols are the factors that various studies have taken into account to verify the usability of these sorts of products and their efficient use in maps [15] [27] [28].

Standardization of tactile symbols is an issue demanded by those involved but at the same time, it is proving controversial, given the difficulty in reaching efficient agreements. On the one hand, there have been efforts as proposed in the International Conference on Mobility Maps in Nottingham in 1972 [29] and, on the other, some countries such as Australia [30], Brazil [31], and Japan [32] among others, have adopted or are in the process of adopting specific standards.

Beyond the difficulties in the standardization of tactile symbols, in this case, some design considerations in its use should be noted. In his manual, Edman talks about point, linear, and superficial symbols. Each of these types is used to present specific information. In mobility plans, point symbols represent specific locations, lines can express direction and guidance, and superficial symbols cover certain areas. All these symbols are informative and can be categorized as 'flat symbols', which are in fact those used most regularly at present.

In summary, and from the point of view of design and usability, the difficulty of identifying symbols in a map increases with the amount included. Therefore, the use of symbols in a map is always the minimum possible: if more than 6 symbols are included a blind person could have trouble memorizing them. The space between symbols should be no less than 3 mm to ensure proper differentiation in relief. It is best not to use similar symbols together as they may be confused, for example, two circles of similar sizes, one outlined, and the other filled in. The minimum size for a symbol to be recognized is around 5 mm. These are just some of the considerations cited by Edman [6].

However, the representation of symbols on conventional tactile maps follows the guidelines of using the three design elements mentioned above, which are part of an essentially two-dimensional nature and are exposed to the sense of touch by the slight elevation.

However, other areas of knowledge such as ergonomics [33], also focused on the study of displays adapted to human use, show that it is possible to use volumetric elements in tasks where one of the requirements is a high degree of tactile discrimination. This is the case with the controls of an airplane which should be distinct and discriminatable to touch in order for pilots to avoid fatal errors. These controls use keypads which the discipline of ergonomics studies from the standpoint of efficiency of use [34].

## III. STUDY DESCRIPTION

### A. Objective and contextualization

The research team planned a series of tests to select a range of easily identifiable and legible volumetric symbols using the sense of touch.

Thus, the main objective of these tests was the selection of a series of three-dimensional symbols easily identifiable using the sense of touch. As a starting point, basic solid figures were chosen by the authors, taking into account that the literature barely deals with these shapes used as tactile symbols. This also followed some earlier studies carried out by the authors suggesting the possibility of this line of research [35] [36]. This is the main innovation in this study.

This selection of symbols will be used in further works in order to determine if this type of symbol can be useful in the configuration of tactile maps.

In contrast, this work focuses only on the study of shape and size factors and does not examine texture and surface quality, resistance, production, cost, or other possible factors that could be of interest in future studies or research.

### B. Methodology

The methodology used for data collection is fundamentally based on the use of tasks with users and prototypes (mock-ups) [37]. In addition, qualitative research techniques and direct observation [38] are used.

### C. Material used in the study

The material used in this study has been produced using 3D printing. In order to improve the surface quality of this system, a thin transparent acrylic layer was applied to the

surface of the tactile symbols. However, it must be emphasized that the study does not attempt to tackle the differences in surface texture, but the differences in shape and size.

Fifteen categories of tactile symbols have been designed and analysed, with eleven of these categories belonging to volumetric elements while the rest are flat elements.

Elementary forms like spheres, cubes, rectangular prisms, cylinders, pyramids, inverted pyramids, cones, inverted cones, and rings (hollow cylinders) have been used in the design of the category of volumetric symbols. As regards the category of flat symbols, the 'U', 'L', 'V' and 'O' shapes, some of these are shown in Fig. 3. The proportion of some simple shapes (cylinder, pyramid, rectangular prism, etc.) used in this study were chosen using direct ratios (1, 1/2, or 1/4, or their respective multiples), as for example, in the case of the rectangular prism,  $1 \times 2 \times 1$ .

In addition, a select group of symbols has been generated using some formal variations, adding complexity to the initial form: angled cut, angled concave cut... in order to assess the users' ability to identify added features using the sense of touch (Figs. 4 and 7).

In addition, to determine the minimum sizes recognizable, each category of symbols has also been represented in a reduction scale of three different dimensions at the base (0.25; 0.5, and 0.75 cm) (Fig. 5). Following Edman's recommendations a standard size of 5 mm was considered [6] (Figs. 5 and 6).

All symbols were coded according to their geometry, size, and additional features, and presented on a white 20x20 mm square base.



Figure 3. Sample of symbols used in the study.



Figure 4. Sample of formal variations to add complexity to some symbols. The picture shows simple shapes on the left (sphere and rectangular prism), incline cut operations in the centre and concave cut transformations on the right.



Figure 5. Sample of reduction scale of pyramid and cube.

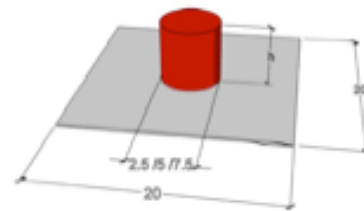


Figure 6. Sample of symbol dimensions on square base (millimeters).

#### D. Test content

Tests consisted of verbal descriptions of the symbols while participants touched the 80 symbols one by one. This assumed that if a symbol is verbally describable, then it will be perceptible and recognizable under other conditions using the sense of touch. The correct answer rates for the 80 symbols were measured.

#### E. Sample and user profile

A sample of 23 participants, 22 blindfolded and one blind user, performed the test. The age range of participants (12 male and 11 female) was from 25 to 55 years old.

#### F. Test procedure

The research team showed participants the symbols one by one, asking them to provide verbal descriptions. The symbols were shown in no particular order until all 80 had been completed. Each symbol was shown on its base. Participants were seated in a comfortable and relaxed position. Users could rotate and manipulate them freely, but not pick them up off the surface of the table.

The description had to be as precise and short as possible. The research team ensured, when beginning the test, that users knew the nomenclature of the geometric shapes in order to refine the description and avoid misunderstandings. In addition, the research team provided participants with a brief introduction to the kind of forms and shapes that they were likely to find, also warning them that some symbols had additional features which participants were to describe if they perceived them.

Participants could use analogies to describe the perceived forms. There was no time limitation, although participants were invited to get the task over with quickly. The research team only intervened with encouraging comments, without providing clues. All the descriptions of symbols which did not adjust to the features were considered errors. Moreover, if participants did not perceive and describe the formal variations of some symbols with additional features, the

research team also considered this an error. The average time spent on the test was around 50 minutes per participant.

G. Record

The test was recorded in summary reports and Excel spreadsheets. Summary reports included some interesting comments for the study. Some details were collected using direct observation, including gestures and individual exploration strategies.

IV. RESULTS

After carrying out the process with all 23 users, the highest percentage of correct answers occurred in single symbols, i.e., those not including special features (more complex and with additional information). Symbols with angled and concave cutting operations generally produced a greater number of errors, while the angled cutting operation had a greater average of correct answers than the angled concave cutting (Fig. 7). In three cases with symbols with formal variations, the cylinder, the ‘U’, and the ‘L’, these reached equal or better results than their corresponding simple versions (Table 1). The concave cut tended to be confused with the angled cut, especially in small-scale symbols. As the graphs for symbols measuring 0.75 cm per side (Fig. 8) and 0.5 cm (Fig. 9) show, it is possible to appreciate a pattern in which the symbols with no formal variation have a greater amount of hits.

Moreover, in the second graph of 0.5 cm, compared with the first graph, it is possible to appreciate a slight decrease in hits due to the loss of tactile acuity resulting from the reduced scale of the symbols. The figures show discontinuities as formal variants were applied only in some cases, as explained above, in order to reduce the number of symbols in the tests or prolonging these excessively.

Thus, as regards the size of the tactile symbols, the general pattern was that a larger size of symbol produced fewer errors of perception (Fig. 10). Some of the symbols tested – the rectangular prism, the cube and the ‘V’ – obtained better results in the small versions than in the larger ones, although the differences were minimal.

The best described volumetric symbols were the pyramid, the ring (100% of correct answers in all sizes), and the thin cylinder, although the cube and the rectangular prism also produced a high index of correct answers. Generally, all of the symbols mentioned were described quickly and spontaneously. Cylinders tended to be confused with cones due to their similar rounded shape, but the reverse was not the case. Pyramids were distinguished from cylinders and cones thanks to the tactile reading strategy of detecting the edges of the pyramid. Symbols ending in a point were spontaneously distinguished when fingertips touched a pointed form. Spheres had a relatively high error rate, and were often described as a cylinder with a rounded head, errors increasing with the reduction in size. Cubes produced some perceptual illusions as when describing them some participants thought they had unequal sides. Inverted shapes, as is the case of inverted cones or pyramids, produced relatively greater numbers of errors than their corresponding forms in their usual positions, although the larger sizes (0.5

and 0.75 cm side) did produce a higher index of hits: 91.30 %. Using the fingertips to access the lower part of these forms with smaller bases, was a key factor in recognizing them.

As regards the flat symbols, all had an optimum success rate. The least number of errors occurred with the circles (‘O’) and the ‘V’ shapes. There was a tendency for ‘U’ to be confused with ‘V’, as one of the lines could not be perceived, or alternatively, in the case of small-scale symbols, with two parallel lines. ‘L’s with equal sides caused perceptual illusions, as some users perceived one side higher than the other.

Finally, through direct observation, some significant hand gestures have been detected while participants were exploring using the sense of touch. An example of this is the strategy of pinching with several fingers to feel the height in order to distinguish the symbol. This occurred when users explored volumetric symbols such as the thin cylinder, pyramid, or cone. This strategy allows greater precision when distinguishing some tactile forms [33]. On the other hand, some forms were even detected spontaneously through feeling the point with their fingertips, as is the case, for example, of pyramids and cones with pointed tops.

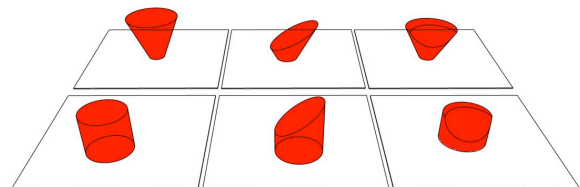


Figure 7. Sample of shape variations of tactile symbols. On the left, simple cylinder or inverted cone shapes; in the centre, shapes with an angled cut; on the right, shapes with an angled concave cut. The last ones obtained the worst rates.

TABLE I. DIFFERENCES OF AVERAGE BETWEEN SIMPLE AND COMPLEX SHAPES FOR SET OF 0.5 CM SYMBOLS

Sample of Tactile Symbols	Correct answer rates (%)		
	Simple shape	Inclined cut	Concave cut
Sphere	91.30	56.52	*
Rectangular prism	91.30	82.61	68.18
Cylinder	91.30	91.30	*
Inverted cone	91.00	*	65.22
L	73.91	82.61	17.39
U	82.61	82.61	26.09
V	95.65	56.52	52.17
<b>Total</b>	<b>88.20</b>	<b>75.36</b>	<b>50.49</b>

\* Not all the variations were tested for all the sizes in order to reduce the duration of the user test to 50 minutes.

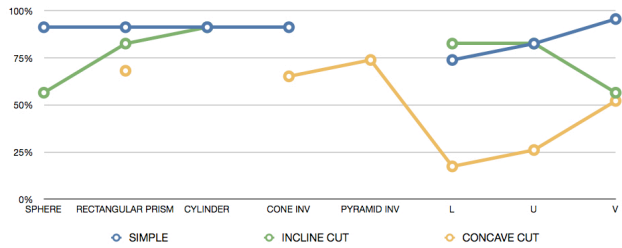


Figure 8. Percentage of correct answers for symbols with formal variations: simple, angled cut, and concave. It shows how simple symbols produced better results.

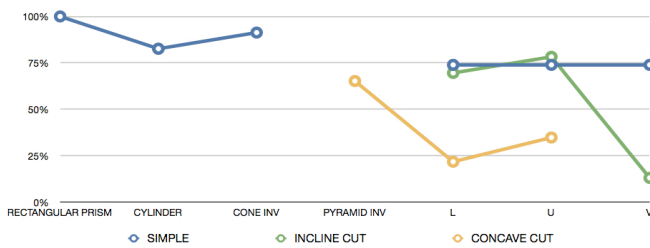


Figure 9. Percentage of correct answers for symbols on different scales. The graph shows a decrease of correct answers compared to figure 8.

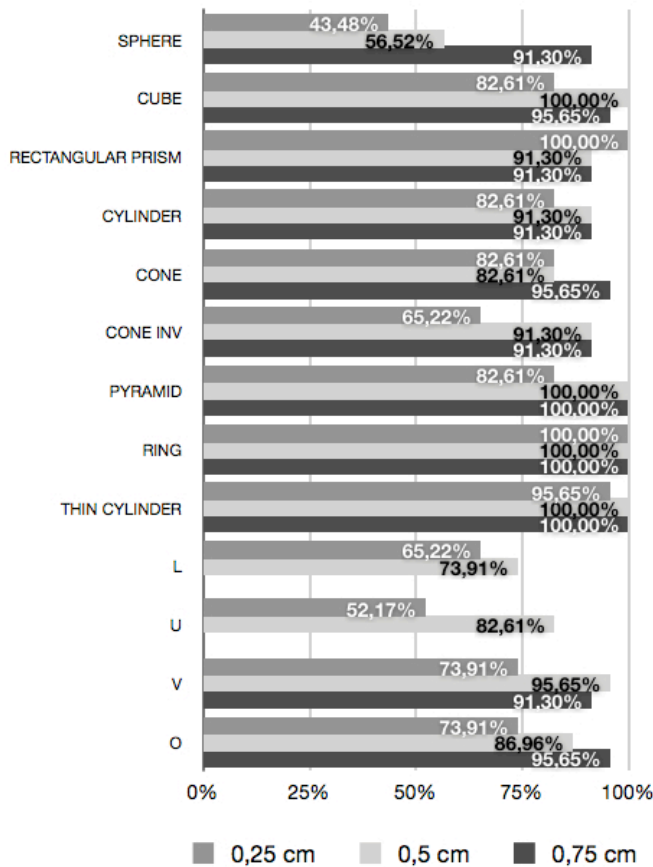


Figure 10. Percentage of correct answers comparing tactile symbols of different sizes.

## V. CONCLUSIONS AND FURTHER WORK

A first conclusion is drawn from the results obtained: given the amount of errors in identifying tactile symbols with formal variations, simple volumetric symbols are more easily perceptible to the sense of touch than complex symbols, thus answering the second research question proposed at the start. This could indicate that, in this case of tactile perception, it seems to fulfil the visual gestalt principle of simplicity.

Furthermore, as expected, larger symbols were more easily identifiable than small symbols. There were no significant variations between the symbols measuring 0.5 cm and 0.75 cm square. Although three of the symbols – the rectangular prism, the cube and the ‘V’ – seem to have an optimum size of less than 0.75 cm, it would be interesting to study this phenomenon in depth in later studies, as certain larger shapes may produce unexpected results outside the general pattern.

However, in the 0.25 cm square symbols, the number of errors increased significantly. This leads us to the conclusion that a size of 0.5 cm square or larger may be appropriate for the design and implementation of three-dimensional symbols in future tactile maps. After analysing the data collected, the best volumetric symbols have been: pyramid, thin cylinder, cube, and rectangular prisms. These presented 100% correct answers in at least one of three sizes.

Regarding two-dimensional symbols, the circle ‘O’ and the ‘V’ achieved optimal hit rates. The flat symbols could be placed on a second level of volumetric accuracy after those mentioned above, but with no significant differences.

Some of the gestures observed during the tests with users, for example, the way in which fingers are used to pinch and explore some volumetric symbols, could be an interesting subject of study.

The fact that some forms have been detected, even spontaneously, thanks to certain formal attributes could open future paths of experimentation in terms of strategies for the exploration of tactile maps. One such example is the effect of pinching fingertips when examining volumetric forms such as pyramids, cylinders, or thin cones. This fact suggests that these sorts of symbols could be used as point symbols. Also, the process for scanning relief maps with volumetric features requires specific strategies that have no place in flat formats.

In addition, certain perceptual illusions detected during the experiment suggest that some curious phenomena, known in detail in the field of visual perception, also play a part in the sense of touch. In future work it would be interesting to study the differences in the proportion of the different solid shapes, for example, the ratios for correctly distinguishing a rectangular prism from a cube.

Regarding the first research question posed, the results of this study suggest that it would be possible to extend the range of discriminatable symbols to include the category of volumetric symbols within the current set of symbols, specifically some basic prisms, which are apparently easily identifiable. Although we must add that it is still early to assert this claim rigorously, and it is necessary to test these symbols in future volumetric tests, for example, in context in real tactile maps with different symbols close to each other,

and not just isolated like those in this study. The optimal distance for correctly distinguishing the set of symbols when these are employed together should also be taken into account.

Thus, the authors assume that these two categories of symbols, 'flat symbols' and 'solid shapes', are difficult to confuse with each other, and would therefore be easy for the users to memorize when used together. Therefore, it would be possible to use more than 6 symbols per key or, at least, to use 6 symbols per key with formal differences that would allow us to avoid discrimination mistakes when using a tactile map. Further studies should deal with the maximum amount of tactile symbols that can be used in a map key to be memorized by users, including those discussed here and those proposed by the authors.

Thus, following the completion of this first trial, further works are pending to clarify the initial. Some of these works have already started in pilot mode using an improved sample of visually impaired persons, showing some initial positive results (Fig. 11).



Figure 11. Pilot study of 3D symbols on relief maps.

#### ACKNOWLEDGEMENTS

The work reported here has been a part of the research project '*Estudio y diseño de elementos de orientación, soportes de comunicación y otros accesorios para la mejora de la accesibilidad en distintos ámbitos de interpretación del patrimonio natural y/o construidos*' supported by the Spanish Ministry of Science and Innovation (project DPI2008-03981/DPI). The authors wish to thank the *Centre de Recursos Educatius (ONCE)* and the *Associació Discapacitat Visual Cataluña B1+B2+B3*, Barcelona, for supporting this research. Finally, this work has been supported by the *Programa de Mobilitat del Personal Investigador de la Universitat Jaume I (E-2010-32)* and the *Fundació Caixa Castelló-Bancaixa*.

#### REFERENCES

[1] AENOR, "UNE-EN ISO 13407, Procesos de Diseño para Sistemas Interactivos Centrados en el Operador Humano", 2000.  
 [2] J. Rowell and S. Ungar, "The World of Touch: an International Survey of Tactile Maps. Part 1: production", *British Journal of Visual Impairment*, vol. 21, 2003, pp. 98-104.  
 [3] C. K. Chua, K. F. Leong, and C. S. Lim, *Rapid Prototyping: Principles and Applications*, New Jersey: World Scientific, 2003.  
 [4] J. L. Clarkson, R. Coleman, S. Keates, and, C. Lebbon, *Inclusive Design: Design for the Whole Population*, Berlín, Germany: Springer-Verlag, 2003.

[5] D. Katz, "Der Aufbau der Tastwelt (The World of Touch). Translated by L. Krueger (1989)", *Der Aufbau der Tastwelt (The World of Touch)*. Translated by L. Krueger (1989), 1925.  
 [6] P. Edman, *Tactile Graphics*, New York American Foundation for the Blind, 1992.  
 [7] J. Lillo-Jover, "Gráficos Tangibles y Orientación en el Invidente", *Psicothema*, vol. 4, 1992, pp. 429-444.  
 [8] M. Blades, S. Ungar, and C. Spencer, "Map Use by Adults with Visual Impairments", *The Professional Geographer*, vol. 51, 2010, pp. 539-553.  
 [9] C. Spencer and J. Travis, "Learning a New Area with and without the Use of Tactile Maps: a Comparative Study", *British Journal of Visual Impairment*, vol. 3, 1985, pp. 5-7.  
 [10] J. Lillo-Jover, "Tacto inteligente: El papel de las Estrategias de Exploración Manual en el Reconocimiento de Objetos Reales", *Anales de Psicología*, vol. 8, 1992, pp. 91-102.  
 [11] C. Perkins and A. Gardiner, "Real World Map Reading Strategies", *The Cartographic Journal*, vol. 40, 2003, pp. 265-268.  
 [12] S. Ballesteros, "Percepción Háptica de Objetos y Patrones Realizados: una Revisión", *Psicothema*, vol. 5, 1993, pp. 311-321.  
 [13] J. Rowell and S. Ungar, "Feeling Our Way: Tactile Map User Requirements- A Survey." *Proc. International Cartographic Conference*, La Coruña. 2005.  
 [14] J. Rowell and S. Ungar, "The World of Touch: an International Survey of Tactile Maps. Part 2: Design", *British Journal of Visual Impairment*, vol. 21, 2003, pp. 105-110.  
 [15] E. P. Berlá, "Haptic Perception of Tangible Graphic Displays", *Tactual Perception: a Sourcebook*, 1982, pp. 364-386.  
 [16] D. McCallum, K. Ahmed, S. Jehoel, S. Dinar, and D. Sheldon, "The Design and Manufacture of Tactile Maps Using an Inkjet Process", *Journal of Engineering Design*, vol. 16, 2005, pp. 525-544.  
 [17] L. Thompson and E. Chronicle, "Beyond Visual Conventions: Rethinking the Design of Tactile Diagrams", *British Journal of Visual Impairment*, vol. 24, 2006, pp. 76-82.  
 [18] A. Voigt and B. Martens. "Development of 3D Tactile Models for the Partially Sighted to Facilitate Spatial Orientation." *Proc. 24th eCAADe Conference (Education and Research in Computer Aided Architectural Design in Europe)*, Bruselas, 2006.  
 [19] L. F. Milan and M. G. C. Celani, "Maquetes Táteis: Infográficos Tridimensionais para Orientação Espacial de Deficientes Visuais", 2008.  
 [20] G. C. Celani and L. F. M. Milan, "Tactile Scale Models: Three-Dimensional Info Graphics for Space Orientation of the Blind and Visually Impaired", *Virtual and Rapid Manufacturing: Advanced Research in Virtual and Rapid Prototyping*, London, UK: Taylor & Francis Group, 2007, pp. 801-805.  
 [21] V- Vovzenilek, M. Kozáková, Z. vSt'ávoová, L. Ludíková, V. Ruuvzivcková, and D. Finková. "3D Printing Technology in Tactile Maps Compiling." *Proc. 24th International Cartographic Conference*. 2009.  
 [22] M. P. Corre-Silva, "Imagen Táctil: una Representación el Mundo", *Universitat de Barcelona*, 2004 (Thesis).  
 [23] R. L. Welsh and B. B. Blasch, *Foundations of orientation and mobility*, American Foundation for the Blind, 1980.  
 [24] Wong, Wucius, *Fundamentos Del Diseño Bi y Tri Dimensional*. Barcelona: G. Gili, 1986.  
 [25] D. A. Lauer and S. Pentak, *Design Basics*, Wadsworth Pub. Co., 2007.  
 [26] R. Rener, "Tactile Cartography: Another View of Tactile Cartographic Symbols", *The Cartographic Journal*, vol. 30, 1993, pp. 195-198.  
 [27] L. L. Lambert and S. L. Lederman, "An Evaluation of the Legibility and Meaningfulness of Potential Map Symbols", *Journal of Visual Impairment & Blindness*, vol. 83(8), 1989, pp. 397-403.

- [28] D. McCallum, S. Ungar, and S. Jehoel, "An Evaluation of Tactile Directional Symbols", *British Journal of Visual Impairment*, vol. 24, 2006, p. 83.
- [29] G. A. James, "Mobility Maps", *Tactual Perception: A Source-Book*, 1982, pp. 334-363.
- [30] A. D .O. N. Mapping, *Symbols for Tactual and Low Vision Town Maps*, [Canberra]: The Division, 1986.
- [31] R.E. Nogueira and B. Florianópolis-Santa Catarina, "Standardization of Tactile Maps in Brazil".
- [32] H. Fujimoto, "Standardization of Display Methods for Tactile Guide Maps for Buildings, Stations and Other Public Facilities." *Proc. Tactile Graphics 2005*, 2005
- [33] S. Pheasant and C. M. Haslegrave, *Bodyspace: Anthropometry, Ergonomics, and the Design of Work*, Boca Raton: Taylor & Francis, 2006.
- [34] M. S. Sanders, *Human Factors in Engineering and Design*, New York: McGraw-Hill, 1993.
- [35] J. Gual, M. Puyuelo,, J. Lloveras, "Universal Design and Visual Impairment: Tactile Products for the Heritage Access". In 18<sup>th</sup> International Conference on Engineering Design (ICED 2011), vol. 5, 2011, pp 155-164.
- [36] J. Gual, M. Puyuelo,, J. Lloveras, "Three-dimensional Tactile Symbols. Relief Maps for the Visually Impaired". *Proc. 4<sup>th</sup> World Conference on Design research, Doctoral Coloquium*, 2011.
- [37] H. Aldersey-Williams, J. Bound, and R Coleman. *The Methods Lab: User Research for Design*. UK: Design for Ageing Network (DAN), 1999.
- [38] B. Laurel, *Design Research: Methods and Perspectives*, Cambridge, Mass.: MIT Press, 2003.

## Ethnographic Examination for Studying Information Sharing Practices in Rural South Africa

Amandeep Dhir<sup>#</sup>, Imad Moukadem<sup>+</sup>, Nobert Jere<sup>@</sup>, Puneet Kaur<sup>#</sup>, Sari Kujala<sup>!</sup>, Antti Ylä-Jääski<sup>#</sup>

<sup>#</sup>Department of Computer Science and Engineering, School of Science, Aalto University, Finland – 02150

<sup>+</sup>Computer Science Department, Gulf University for Science and Technology (GUST), Kuwait

<sup>@</sup>Telkom Center of Excellence in ICTD, Computer Science Department, University of Fort Hare, South Africa, Alice-5700

<sup>!</sup>Department of Design, School of Art and Design, Aalto University, Finland - 00076

{firstname.lastname@aalto.fi}<sup>#!</sup> {moukadem.i@gust.edu.kw}<sup>+</sup> {njere@ufh.ac.za}<sup>@</sup>

**Abstract**— This progress paper describes a study plan and research theme with the focus on information sharing practices, social habits and behavioral access patterns in the context of the mobile phone and ICT usage of rural users in the emerging economies such as South Africa. The aim is to design new solutions to support rural users and their economic existence. An ethnographic field study was performed in order to understand the living conditions of rural communities and the related problems and opportunities in the field of mobile phone and ICT usage. A variety of qualitative measures such as contextual inquiry, participant observation, design exercises, focus and individual interviews were used during the field. However, only preliminary results from field studies are described in this paper. In our study, we found that needs and expectations of the users dwelling in rural context of different emerging economies are very much different from that of developed world. Product designers must consider these needs in order to deliver successful ICT and mobile-based services.

*Keywords*—Emerging markets; ethnography; information sharing; oral culture; rural users.

### I. INTRODUCTION

Telecom sector has witnessed unprecedented growth in the last decade due to which mobile phones have penetrated deep inside our everyday life. Mobile phones have now turned into a commodity item that was considered an element of luxury, a decade ago. Majority of the mobile phones are now owned and used by people dwelling in resource constraint rural communities of emerging markets. Emerging markets are defined as nations holding low to middle income earning people for example India, China and South Africa [1]. According to an estimate, 80% of world's total population is living in these emerging markets and they account for 20% of world's economy [2]. World economists have predicted that emerging markets will continue to show twice growth rates as compared to developed nations [2]. This has resulted in ever-growing interest in exploring and investing in these emerging markets. Technology and related infrastructure have been special area of interest. However, we believe that low income, illiterate and resource constraint communities dwelling in different emerging markets are understudied. Hence, use of Information and Communication Technologies (ICT) and mobile phones in business and daily routines by these resource constraint communities is still a young field of research. Our study is concerned with the examination of

the issues involved in the development of interactive technologies for rural environments. The rural environment studied in this paper is actually the remote region of the development world.

After the invention of mobile phones, scientists considered mobile phones as new mediums for communication and connectivity. But, with the emergence of mobile services such as text messaging and multimedia have transformed this notion. In resource constraint regions of different emerging markets, mobile phones are used for micro-financing and sound medium for carrying out different business activities [3][19]. Mobile phone is now increasingly used for accessing information pertaining to health, education, e-services, government welfare schemes and most importantly farming and related practices. According to a recent estimate, only India and China account for more than 2 billion mobile phone users [4]. Mobile phones are not only portable but also they are easy to use, cost effective and above all, they have become very much affordable. Mobile phones are seen as tools for bridging the existing digital divide between developed and developing world. Previous work on ICT and development argues that under utilization of the ICT infrastructure is one of the main reasons behind this increasing digital divide. We see ICT as an essential catalyst that can potentially help local indigenous markets to expand and develop beyond the local borders.

So far, the users dwelling in resource constraint rural communities of emerging markets are understudied. There are not many studies on developing ICT solutions for resource constrained, low income and digitally unstable communities [17][18][19][20][21][24][25][26][27][28]. We argue that these users are having different needs, expectations and requirements from ICT use. These needs are different from that of users living in developed world. Hence, product developers and ICT designers must understand and examine these different needs in order to manufacture successful ICT adoption in these emerging markets. In our study, we have also emphasized on studying local needs, social habits and behavior patterns of users dwelling in rural communities that are also resource constraint.

The rural areas of emerging markets such as South Africa and India suffer from a lack of economic activity due to which the problem of poverty is further exacerbated.



Rural communities possess a huge potential in developing opportunities for agriculture and other small enterprises, but due to limited or inefficient markets and marginalized basic infrastructure, these communities are not strong enough to tackle poverty and improve rural living. Mobile phone services and ICT possess the required potential for opening these rural markets to the outside world.

In our research, we are seeking to better understand the living conditions of rural residents in India and South Africa so that later, based on the study results, we can propose and develop innovative solutions for improving rural life. Through the ethnography field study we want to identify the existing problems, opportunities and their possible solutions. Our goal of designing new solutions for the rural users in India and South Africa will be achieved through a cross-cultural ethnographic study of rural communities in both these countries. However, in this paper, we have only reported our ethnography field study and its results from two rural communities in South Africa. In this study, we accomplished two main objectives - First, we studied the information-sharing practices in the rural regions of South Africa in order to understand if and how new mobile phone design and ICT solutions can improve their lives. Second, we have cross-examined the findings and understanding present in the existing literature on ICT and development from India and China in context to South Africa. So, in a way, our empirical study is partially cross-cultural by nature.

The rest of the paper is organized as follows. Section II outlines the description of our field sites. Section III briefly describes the related work to our topic. Section IV describes study methodology by explaining the used research methods, study participants and the process of data collection and analysis. Section V presents the themes under investigation in our study and Section VI outlines the results of our ethnography study. Finally, our paper ends with presenting study conclusion and future works.

## II. FIELD SITES

Our study is based on understanding the information sharing practices, ICT use and ownership in resource constraint rural communities of emerging markets where low literate and low income users dwell. In this direction, we have selected two rural but resource constraint communities in South Africa. Our ethnography study is based on the two rural communities located in the Eastern Cape of South Africa, i.e., Alice and Dwesa rural communities. University of Fort Hare is located within Alice while Dwesa rural community is located on the south-eastern coast of South Africa that comprises of an area spread over 235 square Km. The closest town to Dwesa is 50 km away. Dwesa suffers from a low human development index i.e., 0.41 [5], far below the national average of 0.61 [6]. Dwesa also has a low literacy rate at 44.24% against the national rate of 68.5% [7], an unemployment rate of around 78%, and increasing dependence on agriculture, state grants and remittances from urban areas [5]. The working

population of this region is mainly involved in farming related practices, while marginal numbers are dependent on art and craft for their subsistence. Due to the high percentage of youth migration to urban areas, Dwesa is left with mostly elderly people and children below 20 years of age [8][9][10][11]. Dwesa and Alice differ in socio-economic status as Alice is more advanced compared to Dwesa in terms of technology use because the Alice rural community lies in the vicinity of bigger towns and people dwelling here are having higher penetration of ICT and mobile phones compared to Dwesa.

## III. RELATED WORK

Given our focus on the day to day practices in rural settings, we concentrated on three aspects of rural living: oral cultures and their mobile phone usage, existing information sharing practices, and the role of ICT and mobile services in farming and other professions in rural settings. First we referred to the existing work in the Alice and Dwesa rural communities in order to get a basic understanding of the current use and ownership of mobile phones and ICT in these communities [7][8][9][10][11][22]. After this we looked at the existing research in the Human-Computer Interaction for Development (HCID) literature [12]. HCID practitioners have recommended focusing on community-centered design rather than user-centered design when development related issues are the main reasons for the research [12]. Similarly to research on Chinese immigrant workers and ICT [9][10][11] our research also turned to family centered design as the majority of the people in the Dwesa rural community have migrated to urban areas, and they are the ones who provide urban resources to the rural families.

Similar to our argument, i.e., users dwelling in resource constraint communities of emerging markets have different needs, expectations and requirements. Winschiers and Fendler [13] also emphasized this difference in their study as they found western usability notions failed desperately in Nigeria. It was found that western world's notion for evaluating usability i.e., learnability, speed and memorability do not sell as criteria's for usability in Nigeria because Nigerians considered other criteria for considering any software successful. Sukuraman et al. [20] argued that findings from the developed nations are not directly applicable in context to emerging markets due to different needs and requirements. They found that the users in resource constraints communities do not have a pressing need for security and privacy in ICT use unlike developed markets. Sherwani et al. [17] presented the concept of oral and literate culture in order to understand the preference for face-to-face information exchanges in both oral and literate cultures. "Intermediated technology use" is recognized as an important research area in HCID research. Intermediated technology use refers to the situation where users cannot use mobile phone services, ICT or other related technologies on their own due to any one of several reasons like poverty, illiteracy, and lack of accessibility or sufficient skills

required for independent use. In such cases, users often become dependent on others or rely on third parties [18] [19][20]. India has been a testing platform for intermediated technology use [18][19][20][24] and these findings can also be applicable to South Africa for the following reasons: Both countries have a high level of illiteracy and unstable development in infrastructure. Both are also emerging economies that are facing a vast rural and semi-skilled population, poor quality of education, uneven economic development and inefficiency in their local markets. Smyth et al. [25] studied the media sharing through mobile phones in India in order to understand the mobile phone use, ownership and adoption in emerging market context. In this study, we found that if a user is highly motivated then existing hurdles in ICT use and adoption are easily crossed. These hurdles are – cost of Internet use, privacy, complex user interfaces, software virus, legal issues and limited Internet and network connectivity. But users due motivation for entertainment overcomes these challenges. However one of the notable findings from this study was that emerging markets are not yet ready for the advanced ICT development.

In the exiting literature on ICT and development world, there are different research agendas for example; ICT4D, HCI4D and recent of all UbiComp4D, all these platforms have more or less one common objective i.e., focus on the needs and requirements of development world. ICT4D refers to ICT for development where the primary focus is on “how technology can solve challenges faced in global development”. UbiComp4D refers to Ubiquitous computing (UbiComp) for development that strives for the integration of development of resource constraint communities and UbiComp research agendas. Similarly HCI4D stands for Human Computer Interaction (HCI) for development that argues for developing new methods and organizing studies for the development of digitally unstable communities. HCI4D came into limelight as a mechanism for countering the challenges faced in applying traditional HCI research methods when contributing to global IT development. All these three different terms have common agenda i.e., to eradicate poverty from neglected socio-economic contexts [17][18][19][20][21][24][25][26][27][28]. In this direction, some of the notable initiatives are GlobiComp, workshop organized on “*Taking UbiComp beyond developed worlds*” at UbiComp 2009 [26]. “Transnational times” workshop at UbiComp 2010 [27] and workshop titled “*Transnational HCI*” organized at CHI 2011 [28]. All these different initiatives emphasized the need for understanding social, economic and cultural practices in emerging markets for potential ICT development.

#### IV. STUDY METHODOLOGY

The ethnographic study was performed in three different phases where the first phase was exploratory in nature. At this stage, we emphasized gathering information on social habits, access patterns in regard to mobile phone and ICT usage, information sharing and exchanges in rural settings.

In the second phase, the information gained from the first phase was validated using the same research methods and study methodology, but with more systematic sampling of participants. In third phase, we performed a cross-cultural study where findings and results available in the existing literature but from different continent were cross examined with the learning and results obtained through our ethnography study

The first phase of our ethnography study was completed in course of a 5-week ethnography study during May-July 2011 in both Alice and Dwesa Communities. The data collected from the first of the study was analyzed using thematic analysis approach [23]. Based on the study findings, we made some adjustments to our study. Some questions were modified and added in order to know more about the unclear facts. Second study was organized with modified questions between September-October, 2011 in course of 5-week ethnography.

The entire ethnography study was carried twice in the course of 6 months. Three researchers from University of Fort Hare organized and performed the ethnography studies. The researchers made six visits to Dwesa and Alice where they stayed for four to five consecutive days on each visit. They stayed with the local residents and followed a daily routine of visiting common meeting places, social gathering, schools, and adult learning centers, farming sites and shopping places. Before conducting the first phase of this study, we took certain steps to ensure that the results of the study were correct and genuine. For example, three researchers were picked who were already familiar with both communities in order to establish trust and a certain degree of comfort with the study participants. This ensured the easy availability and accessibility of the subjects involved. Second, well-known people from both the communities were taken into confidence so that the subjects participated in our study in a motivated and willing fashion. For example, the respective heads of the communities and principals of the adult learning and training centers located at Dwesa and Alice.

##### A. Research Methods

Mixed method research was practiced during our ethnography study i.e., combination of different qualitative research methods such as contextual inquiry, participant observation, design exercises; focus and individual interviews were used for gaining rich user data. We employed the triangulation research principle in the data collection where findings from every research method complemented the other methods. Tables 1 and 2 present the classification of participants based on gender in different user groups. Table 1 presents the classification in first field study while Table 2 represents second field study.

TABLE I. CLASSIFICATION OF PARTICIPANTS IN FIRST STUDY

User Group	Dwesa	Alice
Youth	11 males 17 females	16 males 13 females
Women	14 females	28 females
Farmers	10 males 2 females	5 males 3 females
Old People	6 males 8 females	11 males 6 females
Teachers/ Workers	3 males 8 females	6 males 8 females

TABLE II. CLASSIFICATION OF PARTICIPANTS IN SECOND STUDY

User Group	Dwesa	Alice
Youth	15 males 15 females	11 males 10 females
Women	15 females	15 females
Farmers	8 males 5 females	8 males 2 females
Old People	7 males 8 females	7 males 7 females
Teachers/ Workers	5 males 10 females	6 males 9 females

**B. Study Participants**

In our study, we tried to involve different stakeholders from the both rural communities. This includes youth (below 25 and not working), women (mostly staying at home, married and from all age groups), farmers (all age groups), older people (above 55 years of age), and professionals involved in other than farming (teachers, craftsmen from all age groups). Our user group contains both illiterate and literate participants who either do or do not possess mobile phones.

In both communities, youth was classified into a primary school group (8-12 years) and a secondary school youth (13-30 years). The youngsters interviewed at Dwesa were between 7-20 years of age due to rural to urban migration, as those above 20 years of age had already left the community. The youngsters interviewed at Alice were more technologically savvy compared to the youngsters at Dwesa. The majority of the interviewed youngsters at Alice had mobile phones while in Dwesa few possessed their own mobile phones. As Alice is closer to towns, there are high numbers of children who are the victims of theft. As a result, parents at Alice preferred their children to have mobile phones so that their children could call them in case of emergency. Furthermore, it was found that the youth at Alice play indoor games such as Sudoku and snakes and ladders; while at Dwesa the children rely on traditional outdoor games in groups.

Farming is more common in Dwesa than Alice, and more farmers were available in Dwesa for this reason. Almost all the interviewed farmers both at Alice and Dwesa

possess mobile phones. In both Alice and Dwesa, we found that a large number of subjects joined in discussions but only few of them were active in answering the interview questions. We cannot restrict or eliminate such people from our discussions as it was against social rules in African society so we allowed everyone to join at their own will. In Tables 1 and 2, we mentioned number of only those users who have participated actively in the discussions. Almost all the focus group discussions organized at Alice were informal in nature because in Alice people are either busy doing their work or travel to the nearest big city to seek employment: hence we met most of our subjects at a community hall except for farmers and older people who were interviewed at farming sites and their homes. Figures 1, 2 and 3 presents an overview of different types of discussions organized by researchers in Alice and Dwesa communities.



Figure 1. Researchers organizing focus group discussions at Alice community



Figure 2. Researchers interviewing youth in a local school



Figure 3. Focus discussions with working user group

### C. Data Collection and Analysis

All the qualitative measures were performed in the native language of that region, for example, in Dwesa and Alice rural communities isiXhosa language is widely spoken and written; so, all our qualitative and design methods were practiced in isiXhosa language. Three researchers from University of Fort Hare are fluent in written and spoken isiXhosa, so all observations and conversations were carried in isiXhosa. The data was collected in the form of audio recordings and written field notes while practicing different qualitative measures in the ethnographic field study. The participating researchers transcribed the recordings first in isiXhosa and later translated them into English. The field researchers crosschecked the translations in order to avoid any incorrect translation of the original text and the associated meanings. The collected data was analyzed using thematic analysis approach where common themes and meanings were obtained after objective and systematic process of data analysis [18]. The whole process ends with higher abstraction levels that were carved from the original data.

## V. THEMES UNDER INVESTIGATION

We have identified three themes for investigation under this study that are listed below:

### A. Oral cultures and Mobile phone usage

Both the Alice and Dwesa rural communities possess large numbers of people who represent oral culture, i.e., illiterate people. We wanted to investigate the different mobile phone features used by oral and literate users. Our two main objectives under this theme are:

1. How do oral users make use of different features in their mobile phones? For example, how do oral users save contacts on their mobile phones? Do they seek help from friends and family? Or have they become familiar with the process of saving contacts in the phone address book? In some of the earlier studies [19], it was found that oral cultures use paper diaries for storing contacts, so we also wanted to examine the possible reasons behind this behavior. In what way do oral users save contacts in their mobile phones and their paper diaries? What is the basis, for instance, for indexing and searching for saved contacts in the paper diaries?
2. Investigating the kind of problems faced by oral users in mobile phone use. Our hypothesis on mobile phone use is that oral users make use of only few mobile phone features. There are many such never used functionalities such as organizing contacts to family, friends and office, the speed dial option as a shortcut for calling someone by pressing a single key, and associating icons or colors with contacts. The question

is do oral users really use such functionalities in their daily routine?

### B. Information sharing practices in rural regions

Under this theme, we are investigating the role of face-to-face or oral information exchange in the daily routine of different user groups. How do different user groups receive the updated information related to the product they want to purchase or sell their produce on the market? We are interested in exploring the access pattern of such information exchanges. We believe that ICT should support the existing information sharing practices that are currently based on social habits. We are also examining the access patterns and different channels through which information flows in rural settings. For example, how do different user groups receive national and international news and updated information on new products they want to purchase. How do they perceive information sharing through computers and Internet? Different state owned TV and radio programs are aired, but what is their overall usefulness to different professions such as farmers, shop owners and small business owners? Furthermore, we are interesting in understanding the role of the different education and health oriented programs owned by the respective state governments for our user group.

### C. Role of Mobile phones and ICT in farming related practices

In this topic, we are studying the role of ICT and mobile phone usage in facilitating different farming related practices such as seed selection, sowing, fertilization and disease management, harvesting and selling agricultural purchase on the market. Our three main objectives under this theme are -

1. Investigating the role of text messages and other mobile phone services in different farming related practices. We want to understand the attitude of farmers to different mobile services, and their associated benefits and risks. How do farmers, for example, sell their produce in market? Do they sell locally or in other big cities? How do they get updated information about their crops? Do they have long-standing relationships with the buyers?
2. Examining the role of state owned infotainment media, i.e., radio and TV programs in the daily routines of different user groups such as information related to markets, to weather forecasts, and also to government policies related to their profession. We are interested in examining "how farmers perceive these informing sharing programs".
3. Other farming specific questions under investigation are different crop diversification programs and their usefulness. Crop diversification programs are promoted by the state government so that farmers can earn more

profits, but do farmers agree and see crop diversification as useful?

## VI. PRILIMINARY RESULTS

Our presented study is still in progress and this paper is also a work in progress contribution; so, we also only outlined some of the preliminary findings here that are crucial findings as per already defined themes of investigations.

### A. *Oral cultures and Information sharing practices*

Similar to the existing studies in emerging markets [15,16], we found that in both rural communities, oral or face-to-face communications are most preferred medium for sharing information. Majority of the interviewed participants in Dwesa community had strong affection for oral way of communication as they consider it to be more relaxing, helps in socializing better and free from miscommunication that frequently happen if communication happens through mobile phones. Information sharing in both rural regions is highly unstructured and people consider technology such as mobile phones and Internet as expensive and un-necessary for delivering non-urgent information for example discussing normal daily routine, gossips and other local news. Youth as well as business professional such as farmers and other employed people excessively use mobile phones. Old people make use of mobile phones mostly for receiving phone call. In both Alice and Dwesa communities, it was found that majority of the people have strong opinion that in future technology can never completely replace the oral or face to face communications. This can be interpreted as people in rural communities are not as addicted as are urban users. Furthermore, technology use is still considered expensive while oral or face-to-face communications are considered effortless.

### B. *Role of Mobile phones and ICT in farming related practices*

In Dwesa community, we found that almost all the interviewed farmers are dissatisfied with the government initiatives related to farming practices. Farmers in Dwesa mentioned that crucial information related to farming practices such as seed selection, sowing, weather forecast, insecticide and pesticide selection could be easily provided using mobile phones. Mobile phones are universally and huge number of farmers possesses feature phones, which support text message facility. However, ICT is underutilized like earlier studies in other emerging countries like China [15][16]. A farmer in Dwesa community is depending upon local school and community centers for getting farming related advice. Some farmers even expressed that main reason for agriculture related losses such as bad yield is due to improper information sharing. In Alice community, farmers are in better position compared to Dwesa because Alice is close to bigger towns so farmers have easy accessibility to farming related equipments unlike Dwesa. In

both the communities, we found that mobile phones are excessively owned and used by the farmers for example, informing neighboring farmers about farming and weather updates, getting updates about the yield price from the local markets. Overall, we got the impression that mobile phones and ICT in general are owned and used in high numbers however; ICT is underutilized due to poor government initiatives. Unlike other researchers [21] who believe that emerging markets are not yet ready for the advanced developed, we hold a different opinion on this matter. In both Alice and Dwesa communities, infrastructure is already present in form of mobile phones; so, any kind of text or voice based services can be easily launched and provided to local community.

## VII. CONCLUSION AND FUTURE WORK

This paper described the design of an ethnographic field study, which aims to investigate mobile phone and ICT usage patterns of rural users in developing countries. The goal is then to identify opportunities for mobile phone and ICT usage based on the results of this study. We believe that bringing ICT to remote environments is growing area of interest in various research communities, and thus our study of the relevant practical and cultural issues is very timely. The paper has been structured in a way that its content not only remains interesting for its readers but also thought provoking. In future, we plan to organize similar ethnographic studies in other emerging markets such as India, China and countries in Middle East. We aim to repeat our existing study in India and China. In this direction, we will locate two similar rural communities as we had in South Africa and apply similar research methods so as to perform cross-cultural analysis of the results. This kind of comparison will enable us in deriving important design drivers that will lead to possible innovative designs for different rural communities that are also resource constraint in nature.

### ACKNOWLEDGMENT

The presented work has been funded and supported by Data communication software (DCS) and Information Systems and Service Engineering (ISSEG) research groups at Department of Computer Science and Engineering, School of Science, Aalto University, Finland. We are thankful to the University of Fort Hare, South Africa for providing necessary logistics required for organizing ethnography studies at Alice and Dwesa rural communities. We are grateful to our fellow researchers Siphe Mhlana and Sikhumbuzo Ngenya for their help during the whole study. We also thank the participants from the Alice and Dwesa communities who took the time to participate in the different phases of this study and share their opinions and experiences. We express our great respect and honor for village heads in Dwesa community who supported our study in every possible way.

REFERENCE

- [1] What Is An Emerging Market Economy? August, 2009. <http://www.investopedia.com/articles/03/073003.asp#axzz1diEas23D> (retrieved: January, 2012)
- [2] Emerging Market Growth Poles are Redefining Global Economic Structure, Says World Bank Report, 17 May, 2011 (retrieved: January, 2012)
- [3] N. Rangaswamy and N. Sambasivan. Cutting Chai, Jugaad, and Here Pheri: towards UbiComp for global community. *Personal Ubiquitous Computing*. 15, 6 (August 2011), pp. 553-564.
- [4] Global mobile statistics 2011, <http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats> (retrieved: January, 2012)
- [5] M. McCann. District Profile Eastern Cape Amathole District Municipality (DC12), Programme of Support to Local Economic Development in the Eastern Cape, Eastern Cape Competitive Advantage Assessment And Training Support Project, European Consultants Organisation, 2006. (retrieved: January, 2012) <http://www.thinasinako.co.za/files/documents/290806175552.pdf>
- [6] UNDP. "Human Development Report 2007/2008 - Country Fact Sheets- South Africa," 2007. (retrieved: January, 2012) [http://hdrstats.undp.org/countries/country\\_fact\\_sheets/cty\\_fs\\_ZAF.html](http://hdrstats.undp.org/countries/country_fact_sheets/cty_fs_ZAF.html)
- [7] Statistics South Africa, Provincial Profile 2004: <http://www.statssa.gov.za/publications/Report-00-91-09/Report-00-91-092004.pdf> (retrieved: January, 2012)
- [8] L. Dalvit, H. Muyingi, A. Terzoli, and M. Thinyane. The Deployment of an e-Commerce Platform and Related Projects in a Rural Area in South Africa, *International journal of Computing and ICT Research*, Vol. 1, No. 1, pp. 9-18.
- [9] C. Pade-Khene, R. Palmer, and M. Kavha. A baseline study of a Dwesa rural community for the Siyakhula Information and Communication Technology for Development project: understanding the reality on the ground. *Information Development*. Information Development Journal November 2010 vol. 26 no. 4, pp. 265-288.
- [10] G. N. Sicelo, Implementation of a Virtual shopping mall for Dwesa, a rural area in the Eastern Cape, South Africa. MSC thesis, Department of Computer Science, University of Fort Hare, South Africa.
- [11] Dwesa Village Connection Business Modelling, Feasibility Analysis, October 2008 (retrieved: January, 2012) [http://www.ungana-afrika.org/sites/default/files/documents/dwesa\\_business\\_modelling\\_report.pdf](http://www.ungana-afrika.org/sites/default/files/documents/dwesa_business_modelling_report.pdf)
- [12] A. Dearden, User-centered design considered harmful. *Information Technologies and International Development* 4, 3 (2008), pp. 7-12.
- [13] M. Friedewald and O. Raabe, Ubiquitous computing: An overview of technology impacts, *Telematics and Informatics*, Volume 28, Issue 2, May 2011, pp. 55-65.
- [14] X. Lang, E. Oreglia, and S. Thomas. Social practices and mobile phone use of young migrant workers. In *Proc. MobileHCI2010*, ACM Press (2010), pp. 59-62.
- [15] X. Liu, J. Liu, J. Cai, Y. Liu, and X. Wang. Design for China's migrant workers: A case of user research and mobile product concepts development. In *Proc. HCD09*, Springer-Verlag, Berlin, Heidelberg, pp. 482-491.
- [16] E. Oreglia, Y. Liu, and W. Zhao. Designing for emerging rural users: experiences from China. In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*. ACM, New York, NY, USA, pp. 1433-1436.
- [17] J. Sherwani, N. Ali, C. P. Rosé and R. Rosenfeld, Orality-Grounded HCID: Understanding the Oral User, *Information Technologies and International Development*, vol. 5, Dec 16 2009, pp. 37-49
- [18] T. Parikh and K. Ghosh. Understanding and designing for intermediated information tasks in India. *IEEE Pervasive Computing* 5, 2 (2006), pp. 32-39.
- [19] N. Sambasivan, E. Cutrell, K. Toyama, and B. Nardi. Intermediated technology use in developing communities. In *Proc. CHI2010*, ACM Press (2010), pp. 2583-2592.
- [20] A. Sukumaran, S. Ramlal, E. Ophir, V. R. Kumar, G. Mishra, V. Evers, V. Balaji, and C. Nass. Intermediated technology interaction in rural contexts. In *Ext Abstracts CHI 2009*, ACM Press (2009), pp. 3817-3822.
- [21] E. Brewer, M. Demmer, M. Ho, R. J. Honicky, J. Pal, M. Plaqueche, and S. Surana. The challenges of technology research for developing regions. *IEEE Pervasive Computing* 5, 2 (2006), pp.15-23.
- [22] E. A. Emmanuel and H. N. Muyingi. A mobile commerce application for rural economy development: a case study for Dwesa. In *Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT '10)*. ACM, New York, NY, USA, pp. 58-66.
- [23] L. Rourke, T. Anderson, D. R. Garrison, and W. Archer. Methodological Issues in the Content Analysis of Computer Conference Transcripts. *International Journal of Artificial Intelligence in Education* 12 (2000), pp. 8-22.
- [24] A. Joshi, N. Welankar, N. BL, K. Kanitkar, and R. Sheikh, Rangoli: a visual phonebook for low-literate users. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services (MobileHCI '08)*. ACM, New York, NY, USA, pp. 217-223.
- [25] T. N. Smyth, S. Kumar, I. Medhi, and K. Toyama. Where there's a will there's a way: mobile media sharing in urban india. In *Proceedings of the 28th international conference on Human factors in computing systems (CHI '10)*. ACM, New York, NY, USA, pp. 753-762.
- [26] Globicomp: Taking UbiComp beyond developed worlds. <http://www.cs.swan.ac.uk/globicomp2009/>, UbiComp 2009 Workshop, 30 September, 2009 (retrieved: January, 2012)
- [27] I. Shklovski, S. Lindtner, J. Vertesi, and P. Dourish, Transnational Times: Locality, Globality and Mobility in Technology Design and Use. *Ext. Abst. UbiComp 2010*, ACM Press (2010), pp. 515-518.
- [28] J. Vertesi, S. Lindtner, and I. Shklovski, Transnational HCI: humans, computers, and interactions in transnational contexts. *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems (CHI EA '11)*. ACM, New York, USA, pp. 61-64.

## Accessibility Study of Rich Web Interface Components

Juliana Cristina Braga  
Universidade Federal do ABC  
Santo André, Brazil  
e-mail: juliana.braga@ufabc.edu.br

Rodrigo Torres Leme  
Universidade Federal do ABC  
Santo André, Brazil  
e-mail: rodrigo.torresleme@gmail.com

Rafael Jeferson Pezzuto Damaceno  
Universidade Federal do ABC  
Santo André, Brazil  
e-mail: rafael.damaceno@aluno.ufabc.edu.br

Silvia Dotta  
Universidade Federal do ABC  
Santo André, Brazil  
e-mail: silvia.dotta@ufabc.edu.br

**Abstract** – The use of the latest technologies to develop rich interfaces for websites could decrease your accessibility, making problematic the access for people with disabilities. Faced with this problem, this paper executed a study to evaluate the accessibility of rich components on the Web. The paper used WAI-ARIA recommendations as reference and analyzed the accessibility problems found in the Jboss Richfaces components library. The study demonstrates a methodology to make the analyzed rich components accessible.

**Keywords** – Rich Internet; accessibility; WAI-ARIA; Web 2.0.

### I. INTRODUCTION

Accessibility is defined as the possibility and condition range, perception, and understanding for safely and autonomously use of buildings, space, furniture, urban equipment and elements [1]. The term indicates the possibility of any person to benefit from a life in society, including the internet.

Accessibility on the internet or accessibility on the web means to allow web access for everyone, not depending on the user type, situation or tool. It is concerned with creating or rendering tools and web pages accessible to a larger number of users, including people with disabilities.

The web is a resource that has become more important in many life aspects: education, work, government, commerce, health, leisure, and many more. It is essential that the web is accessible in order to allow people with disabilities to have equal access and opportunities. Also, an accessible web may help people with disabilities to be active part of the society.

However, with the changes in the sites development technologies, accessibility on the web has decreased, and could contribute to increase digital exclusion and to negatively impact the life of the disabled who already uses the internet.

Rich Internet Applications (RIAs) are web applications that employ advanced technologies of user interface to try to bring the interface of web systems closer to the interface of *Desktop* systems [2].

The earliest web applications supported basic HTML only, and although they could provide simple functions,

these applications did not have nor provided the experience of a *Desktop* application.

The general objective of this research is to study solutions which may reduce the negative impacts that new rich technologies induce on the accessibility on the Web. Specifically, the objective of this paper is to study the accessibility of some components of the *Jboss Richfaces* toolkit.

This paper is organized as follows. Section 2 introduces fundamental concepts to understand accessibility evaluation process of a website using rich components. Section 3 shows all technical items and steps that led this study. Sections 4 to 6 describe encountered problems and their resolution, added to the system validation. Section 7 presents the conclusions of this work.

### II. THEORETICAL BACKGROUND

#### A. RIAs Technologies

There are two technologies that facilitate the RIAs development: *toolkits* and *frameworks*. *Toolkits* assist on abstracting the differences of browsers and on providing basic functions to handle events. *Frameworks* assist on standardizing sites development. Some *toolkits* and *frameworks* examples are: Dojo, Web Toolkit Google, Dojo Abra Rico, Prototype, User Interface Library, Zimbra Kabuki AJAX Toolkit, Yahoo User Interface, Jboss Richfaces, JavaFX, Ruby on Rails, and Java Server Faces (JSF).

This paper studies the *Jboss Richfaces 3.3.3.Final toolkit* and the *Java Server Faces 1.2 framework*, based on the World Wide Web Consortium (W3C) recommendations [3].

#### B. World Wide Web Consortium (W3C)

The *World Wide Web Consortium* (W3C) is an international consortium created in 1994, and it is responsible for establishing web standards. More than 400 organizations are members of this Consortium.

The W3C dealt with the accessibility problem in 1998 when the W3C Web Accessibility Initiative (WAI) was launched. Therefore, WAI is the part of W3C responsible for web accessibility.

WAI works with organizations from all over the world to develop strategies, guidelines and resources aiming to make the Web accessible to people with disabilities. One of the WAI roles is to develop guidelines and techniques that describe accessibility solutions for Web software and developers.

Until 2008 WAI did not handle accessibility problems in RIAs and from that year on it started to handle these problems by releasing the *Accessible Rich Internet Applications Suite* (WAI-ARIA 1.0).

### C. WAI-ARIA

WAI-ARIA 1.0 is a set of technical recommendations that is being developed to specify ways to make the components of the rich internet accessible.

In fact, these recommendations suggest the insertion, in the sites source code, of tags of four types: *landmark*, *role*, *state* and *properties* [3].

*Landmark* is an element that indicates the *layout* divisions of a site, and can be header type, navigation bar, footer, and text body, among others [3].

*Role* designates the role executed by a RIA component [3]. As an example, a *slider* component, such as the one presented in Fig. 1, can be created from images and source code written in *JavaScript* language and can include the WAI-ARIA tags to make it accessible for a screen reader.

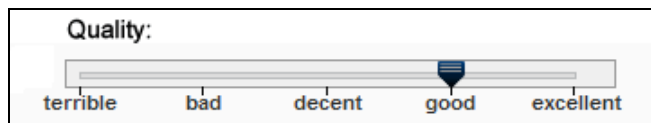


Fig. 1. Slider component created from images and JavaScript source code [3].

*Properties* indicate the property that a given component has [3]. In the example from Fig. 3, the “maximum-value” property could exist. *State* indicates the state of this property; in this example, the “maximum-value” *property* could have the “good” *state*.

A set of WAI-ARIA tags of *live* type also exists, and they are important to inform the screen reader about the updates without full page changes, which are common in sites that use RIAs [3]. An example is the decreasing count of remaining characters to be typed in a text box, a component that is very abundant in social networking services sites. Inserting the tag in this case would allow the screen reader to warn the user about how many characters are left to be typed.

### D. Tools to Evaluate Accessibility

The evaluation of a given site accessibility can be done manually and/or automatically.

Manual evaluation consists of inspecting the site pages source code, in order to analyze its contents based on the W3C guidelines, searching for accessibility errors.

Automatic evaluation consists of testing the pages of a site using evaluation software that detects the code and analyzes their contents. Usually, the software is based on W3C guidelines, and evaluates the accessibility level

inside these set of rules to produce detailed reports automatically. Generally, these tools/reports present the errors and suggestions on how to fix them, as well as the verifications that must be executed manually.

There are dozens of evaluators/validators of websites accessibility available in the market, such as: Cynthia Says [4], Hera [5], Ocawa [6], and W3C Validator [7].

### E. Assistive Technologies

Inside Computer Science scope, assistive technologies are a set of software and hardware designed specifically to help people with disabilities to perform their daily activities [8].

That research studied voice synthesis assistive technology, also known as screen reader. This kind of computer program can reproduce audio of user interactions with the operational system and with others programs, like web browsers. Currently, there are many screen readers, such as: NonVisual Desktop Access (NVDA) [9], Jaws for Windows Screen Reading [10], Virtual Vision [11] and Window-Eyes [12].

## III. STUDY ORGANIZATION

This project was developed in four steps, with some of them being concurrent. The first step was the survey and the bibliographic study already mentioned in the theoretical reference. The second step was the analysis of rich components accessibility; the third step was the adjustment of the rich components accessibility; and the fourth step was the validation of the study components accessibility. The following topics provide details of steps two to four.

### IV. IDENTIFICATION OF THE ACCESSIBILITY PROBLEMS OF RICH COMPONENTS

This study analyses an evaluation software used in a university to measure teachers and disciplines performances. It uses the following rich components: *commandButton*, *commandLink*, *status*, *ajaxValidator*, *column*, *dataTable*, *message*, *messages*, *modalPanel* and *panel*.

The system consists of four pages: access page, with login and password; list of disciplines page, grading page and conclusion page. The source code of each page that uses the components is composed by four parts: HTML, Java Server Pages, Asynchronous JavaScript and XML (AJAX) and JBoss RichFaces.





Fig. 2. The use of JBoss RichFaces components in the evaluation system under study. Access page. 2011.

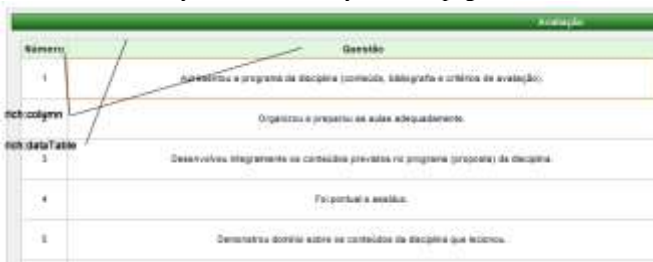


Fig. 3. Use of JBoss RichFaces components in the evaluation system under study. Evaluation page. 2011.

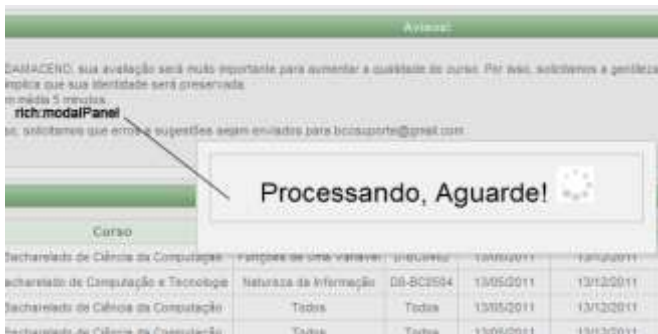


Fig. 4. Use of JBoss RichFaces components in the evaluation system under study. List of disciplines page. 2011.

It is known that an accessible site must firstly be in accordance with the HTML standards suggested by W3C. Therefore, in order to help the identification of the accessibility problems of the study components, an automatic tool for validation of sites suitability to W3C standards was used. The chosen tool was *Validator* from W3C [7], which is free software and does not require installation. This tool was selected due to being the only one in the category that supports WAI-ARIA tags, which are included in the Document Type Definition (DTD) file that defines the set of HTML tags the browser should recognize and interpret. It should be noted that there are other tools available for validation, such as the “Avaliador DaSilva” (Evaluator DaSilva) [13], and “Avaliador e Simulador de Acessibilidade de Sites” – ASES (Evaluator and Simulator of Accessibility of Sites) [14], which do not include (until now) the tags related to WAI-ARIA, making impossible any analysis.

For validation, the “*Validate by direct input*” option was used and it is available through the W3C Validator site. The source code of each opened page in the Mozilla Firefox version 5.0 browser was copied to the validator code field. When the “*check*” is clicked on, the validator reported the problems found. After the analysis result, the validation system detected eight errors on the start page, four on the menu page, and three on the main use page. Some errors are not related to accessibility, and are related to not complying with W3C standards. The *Validator* does not cover all cases and, therefore, it was also necessary to open the source code of the evaluated system on a true text editor.

After the automatic validation and the code inspection, a scan of the system was executed using the NVDA screen reader. The reader helped to detect accessibility problems that could be corrected in the next step of the research. Many problems were detected since the study system was initially developed with no consideration about accessibility. The table I below describes the most common problems found:

TABLE I. RICH COMPONENTS AND PROBLEMS FOUND.

Component	Problem
rich:column	Cells of the table inaccessible via the keyboard. The screen reader does not read its contents.
rich:dataTable	
rich:message	Message generated by AJAX technology is not read by screen reader.
rich:modalPanel	Component inaccessible via the keyboard. The screen reader does not read its contents.
rich:panel	Component inaccessible via the keyboard.

Besides problems found in the rich components, problems in the use of not rich components were also identified, as listed below:

- Use of classic HTML tables for layout;
- Tab order not defined;
- Reading all content on the web pages without pause;
- Reading all the cells of the tables without pause;
- Table inaccessible via the keyboard;
- List items inaccessible via the keyboard;
- Selectable Radio buttons, but with limited access, since it is impossible to read the contents of rich:dataTable cells and to obtain information on the subject.

## V. ADJUSTMENT OF THE SYSTEM ACCESSIBILITY

Based on the accessibility errors indicated by *Validator* and on the problems identified by the screen reader, accessibility adjustment was executed. Initially, the system was adapted to the basic accessibility recommendations WCAG 1 and WCAG 2, such as the deletion of tables used as layout and the correction of tags, both from the HTML part. Then, the WAI-ARIA attributes were inserted in the four pages, inside the HTML tags. It was necessary to change the *Document Type Definition* (DTD), a file that defines the valid

HTML blocks, to a version that recognizes those attributes and is specified by the W3C. *Role*, *aria-describedby* and *aria-live* attributes were inserted inside the HTML tags. *Jboss Richfaces* components do not recognize these attributes, because they were not built with WAI-ARIA support. Therefore, in order to use them, it was necessary to change the source code and compile a new version of the *Jboss Richfaces* toolkit, in which the *aria-describedby*, *role* and *tabindex* were inserted in seven of the components used in the system. *Jboss Richfaces* source code was changed in two ways: through a project manager software based on the *Project Object Model* (POM) concept, with the modification of files (usually with \*.xml and \*.jspx extensions) in which new attributes and insertion tags were inserted, respectively; and manually by opening the source code of each class of a given component and modifying/creating whatever was necessary.

TABLE II. RICH COMPONENTS, PROBLEMS FOUND AND SOLUTIONS

Component	Problem	Solution
rich:column	Cells of the table inaccessible via the keyboard	Putting “tabindex” attribute in each cell of the table to make them accessible.
rich:dataTable	Cells of the inaccessible via the keyboard	
rich:message	Message generated by AJAX technology is not read by screen reader	Putting HTML attribute “aria-live=assertive” inside div element responsible for displaying rich:message.
rich:modalPanel	Component inaccessible via the keyboard	This component is shown to user when he clicks to move from one page to another in order to inform him that the system is loading the next page. It was not modified, yet.
rich:panel	Component inaccessible via the keyboard	This component is used as layout organizer, so, it is not necessary to receive keyboard focus (in the studied system). But, <i>Jboss Richfaces</i> library was modified to recognize the “tabindex” attribute.

The *dataTable* rich component (table with AJAX support) was manually modified at the class in Java language that creates its cells (*cellRenderer.java*) to add the *role* and *tabindex* attributes, with values “*gridcell*” and “-1”, respectively. The first pair (*role*=“*gridcell*”) indicates the role (table cell) and the second one (*tabindex*=“-1”) makes it capable of receiving focus. On the other hand, an alternative to the need to modify the source code of this library and to compile a new version is to insert the attributes through the *JavaScript* language. In the same *JRF* *dataTable* component, this language was used to insert the *aria-describedby* attribute in each table cell, with the designation of the column header it belongs to. In addition, the source code developed in *JavaScript* was created to make possible the navigation on the tables

using the direction keys 'left', 'up', 'right' and 'down', as well as to access the pages evaluation links and the radio button selection items using the 'enter' and 'space' keys. In a test using other browsers (Google Chrome, Opera, Internet Explorer) it was noticed that the source code has different behavior. Adjustments are necessary for the same operation in all possible browsers/screen readers. From the use of a screen reader to open the study system, it was found that words are spoken by a speaker with English language accent, which hinders the reading understanding of the site elements for users that are not proficient in this language. However, it is possible to use an external speaker, built in Brazilian Portuguese language, which improved the understanding of the spoken words by this research test subjects.

## VI. STUDY VALIDATION

The system was tested by subjects with some visual disability. To increase the number of tests, subjects without disabilities also tested the system blindfold. All tests were executed using the NVDA version 2011.2 screen reader (a free software) and the Mozilla Firefox version 5.0 browser. It was noticed at this stage that because the screen reader was built before the arrival of RIAs, its use was difficult and it could not identify all the rich components. However, the NVDA has been constantly updated in order to support the WAI-ARIA tags, which may become accessible components.

## VII. CONCLUSIONS AND FUTURE PROJECTS

If on one side, the technologies to develop rich interfaces for web sites advance, on the other side, the frequent use of these technologies contributes to decrease the accessibility of *Websites*. Faced with this problem, this paper performed a study to evaluate the accessibility of Web rich interface components. The paper analyses and corrects accessibility problems found in a system that uses the *Jboss Richfaces* components library. The study demonstrates how it was possible to make the rich components of the *Jboss Richfaces 3.3.3.Final toolkit* accessible using only WAI-ARIA tags. To insert these tags, it is necessary to modify the *toolkit* source code (and to compile a new one). But, these tags could also be inserted using *JavaScript* language. Since the WAI-ARIA recommendations are generic, that is, they are not applicable to only a group of specific rich components; it is necessary to test these recommendations on other toolkits rich components (like Google Web Toolkit), in addition to execute tests with different browsers/screen readers. In summary, the study corroborates that it is possible to combine sites with rich interfaces and accessibility. However, currently this combination depends on the developer, limiting the dissemination of accessibility on the Web. This paper suggestion is for rich components to be made available to the developer in accordance with WAI-ARIA recommendations. This study presents a methodology to execute this task and it is the main contribution of the paper.

## VIII. REFERENCES

- [1] Brazil. Brazilian Association of Technical Standards (ABNT). Brazilian Standard ABNT NBR 9050: Accessibility of buildings, furniture, equipment and urban spaces. 2nd Edition, 2004. Available at: <http://www.mpdft.gov.br/sicorde/NBR9050-31052004.pdf>. Accessed on: 29 Nov 2011.
- [2] P. Deitel and H. Deitel. Ajax, rich internet applications and web development for programmers. Ed. Pearson. 2008.
- [3] Web Accessibility Initiative – Accessible Rich Internet Applications Suite (WAI-ARIA). Available at: <http://www.w3.org/WAI/intro/aria.php>. Accessed on: 29 Nov 2011.
- [4] Cynthia Says. Available at: <http://www.cynthiasays.com/> Accessed on: 29 Nov 2011.
- [5] Hera. Available on: <http://www.sidar.org/hera/index.php.en> Accessed on: 29 Nov 2011
- [6] Ocawa. Available at: <http://www.ocawa.com/en/Test-your-Web-Site.htm> Accessed on: 29 Nov 2011.
- [7] W3C. Markup Validation Service. Available at: <http://validator.w3.org/>. Accessed on: 29 Nov 2011.
- [8] L. H. Tonet and D. P. Andrés. Research of tools of computer accessibility for visually impaired and W3C Recommendations. 2006. Available at: <http://guaiba.ulbra.tche.br/pesquisa/2006/artigos/sistemas/161.pdf> Accessed on: 29 Nov 2011.
- [9] NonVisual Desktop Access (NVDA). Available at: <http://www.nvda-project.org/>. Accessed on: 29 Nov 2011.
- [10] Jaws for Windows Screen Reading. Available at: <http://www.freedomscientific.com/products/fs/jaws-product-page.asp> Access on: 29 Nov 2011.
- [11] Virtual Vision. Available at: <http://www.virtualvision.com.br/index.html> Accessed on: 29 Nov 2011.
- [12] Window-Eyes. Available at: <http://www.gwmicro.com/window-eyes/> Accessed on: 29 Nov 2011.
- [13] Brazil Accessibility. DaSilva Evaluator. Available at: <http://www.dasilva.org.br/?blogid=2>. Accessed on: 29 Nov 2011.
- [14] Brazil. Ministry of Planning, Budget and Management. Evaluator and Simulator Accessibility of Sites. Available at: <http://www.governoeletronico.gov.br/acoes-e-projetos/e-MAG/ases-avaliador-e-simulador-de-acessibilidade-sitios>. Accessed on: 29 Nov 2011.

## Virtualization Technology for Multi-display Systems

Igor Petukhov, Lyudmila Steshina, Ilya Tanryverdiev

Mari State Technical University, Department of Radio Technical  
Yoshkar-Ola, Russian Federation

Emails: {[PetuhovIV@marstu.net](mailto:PetuhovIV@marstu.net), [SteshinaLA@marstu.net](mailto:SteshinaLA@marstu.net), [ili.tanryverdiev@gmail.com](mailto:ili.tanryverdiev@gmail.com)}

**Abstract** – The method for creation of multi-user multi-monitor systems designed for educational purposes is suggested. The method utilizes virtual machine technology which makes possible to create completely isolated user working environment, remote access to equipment and remote workstation administration. Standard interfaces like DVI, VGA or USB are used as data transmission channels along with network technologies. Such architecture facilitates the operator's work, making it more comfortable, and improves the quality and capabilities of the distant access to remote resources.

**Keywords**–multi-user multi-monitor systems; virtual machine technology; interface.

### I. INTRODUCTION

Intense progress in remote educational techniques and reformation of the existing intramural ones stimulates the information systems developers to seek for newer technologies to be convenient and beneficial for the educational process. First, it concerns the reformation of computerized student workstations organization, namely, their graphics architecture, ways of workstations interaction with each other and with the teacher's workstation. Nowadays, tablet PCs, graphics systems having two or more displays, multi-user designs on the base of a single computer, etc. are being implemented together with non-removable computerized student workstations.

Graphical systems having two or more displays are well known and successfully used by specialists in different spheres, especially where the human-operator has to deal with great amount of changing data. Such systems are used at stock exchanges, in monitoring different natural processes, in the systems of technological process automated control at industrial enterprises and in many other fields.

Most of these systems are designed for the work with a single operator, i.e. they create common information space including a single operator. There are works on creating graphical systems with intelligent terminals and direct allocation of data processing function to several computing machines [1], systems with intelligent satellites [2], client/server models 'host computer plus user terminals' using different types of PCs [3, 4].

Software programs like Citrix providing terminal access for Windows and UNIX and terminal services incorporated in server versions of Windows [5] are among the primitive mechanisms of terminal access known today.

Heterogeneous distributed computer systems for decentralized machinery control using Web-server and

remote web-client servers [6], multi-user system Multiseat implementation technologies [7], corporate solutions of distributed user systems in virtual environment on the base of Microsoft RemoteFX [8] are also of interest.

Most of these systems are designed for the work with a single operator, i.e. they create common information space including a single operator. However utilizing multi-display systems in education may be still more efficient, if their own working environments are created for them.

Moreover, implementation of distributed terminal systems gives an advantage of providing substantial financial saving on purchase of equipment, diminishing operational costs for system maintenance and, in a way, realizing 'green' energy saving concept [9, 10].

Thus, the fact of practicability of multi-user terminal systems implementation is becoming evident.

Meanwhile there still exist a number of problems requiring special attention. They are the problems of providing the system processing speed and its productivity, the security of information communication when transmitting it through the open access networks, implementation of the 'severe real time' mode.

The aim of the given work is to elaborate the way of creating a multi-user multi-monitor system with remote access capability designed to be used as information environment in education process.

### II. THEORETICAL ANALYSIS

Multi-user multi-monitor systems are those built on the base of a single computer and having many monitors each meant to be used by one operator.

In information systems, the so-called terminal principle is usually used to provide multi-user work, interaction of technical provision of operators' workstations through local processing networks or other network technologies underlying the principle.

This causes substantial delays in information communication and limits the communication bandwidth between operators. Technical provision of operators' workstations is common user PCs having input/output devices for interaction with an operator and network adapter for adding a client to the network. Organization of user workstations without using the client PCs has limited implementation so far because of its low dynamic features. For example, ArtistaNET controller making possible to integrate LCD displays into a local area computer network without a client PC has for the time being the productivity [11] shown in the table below.

TABLE I. COMPARATIVE FEATURES OF ARTISTANET CONTROLLER PRODUCTIVITY

Resolution	Interface	Image Transfer Rate
VGA	100 Mbps	16 images per second
SVGA	100 Mbps	12 images per second
XGA	100 Mbps	5 images per second
WXGA	100 Mbps	4 images per second
SXGA	100 Mbps	3 images per second

Such communication bandwidth of data channels between the operators' workstations is evidently insufficient for monitoring the quickly changing processes and creating the 'real time' for a user.

There is another principle, suggested by the authors earlier and based on utilizing analog channels for data transfer. The essence of the technique is in setting a lot of video cards into the computer and connecting each of them with two monitors by a standard cable through DVI or VGA interface (Fig. 1) [12].

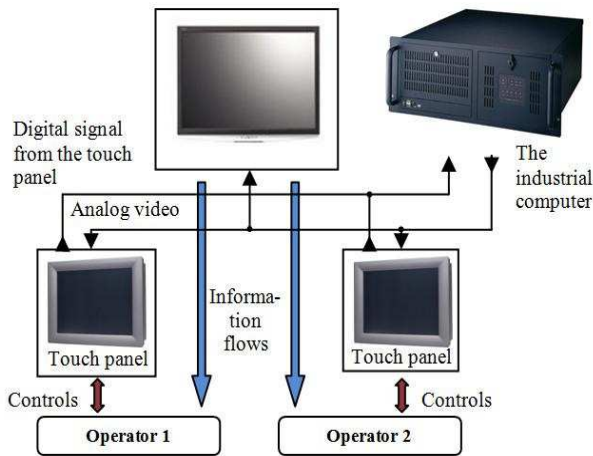


Figure 1. Scheme of separation of information and control functions

Video data are transferred through the cable without any transformation and come directly to the monitor, which completely excludes delays in the channel of data transfer. From the view point of hardware, the method can be used to create small area multi-user information systems, since usage of even special shielded cable to connect the monitors on the operators' workstations cannot provide total noise absence and, as a result, makes enhancing the coverage area infeasible.

An important issue in using the method discussed is software support of isolated users work environment. Every operator usually performs his/her function that is loosely coupled with the function of some other operator. In this case, each operator requires specialized software or hardware

at his/her workstation, e.g. access to the CAD-system being studied. Moreover, such a technology does not make it possible to bring tablet PCs into the information environment to create mobile computer workstations.

Another burning problem appearing in creation of information environment for educational purposes is the problem of multiple access to hardware located in the university premises. It is common that demonstration of this or that effect or phenomenon in educational process involves complex technical means connected to a computer, be it an electronic microscope or some other specialized equipment. As a rule, the existing information systems used in distance learning are not able to supply a remote student with access to the equipment set in the university premises, limiting this by the access only to electronic resources. This results in the fact that remote students are unable to fulfill practical, laboratory and research work at distance and sometimes causes deterioration of distance education quality.

### III. VIRTUAL MACHINE TECHNOLOGY

The analyses of the problems mentioned gives evidence that most of them depend on the implementation of network technologies, namely, local computer networks (LCN). Thus, the solution of the larger part of these can be provided via rejection of LCN usage and implementation of a terminal system through a new technology. According to this, the authors developed an alternative technology the essence of which is in creation of multi-user computational system based on a single computer. To identify such a system it is suggested to use the term 'pseudo terminal system' because of the absence of exteriorized terminals [13].

It is suggested that solving the problem of creating individual isolated environments in multi-user multi-monitor systems should be realized applying virtual machine technology.

A unique virtual machine (virtual terminal) corresponds to each of a number of operators. The terminal has guest operating system and its unique input (keyboard, mouse, touch panel, etc.) and output (monitor, VR-headset, etc.) devices, as well as video adapter and information input interface controllers, as shown in Fig.2.

Virtual machine monitor provides access of guest operating system of each operator to the hardware appointed to it (video adapter, input controller), while guest operating system works with the hardware offered via standard drivers.

In this case, guest operating system of each operator forms a desktop and processes graphical elements in video processor of the appointed to it adaptor, puts the graphical information out by means of the interface of the same physical adaptor without applying virtualization of the device by the monitor of virtual machines.

The information designed to be displayed for the user comes from the interface to the user monitor connected with the video adapter corresponding to this user.

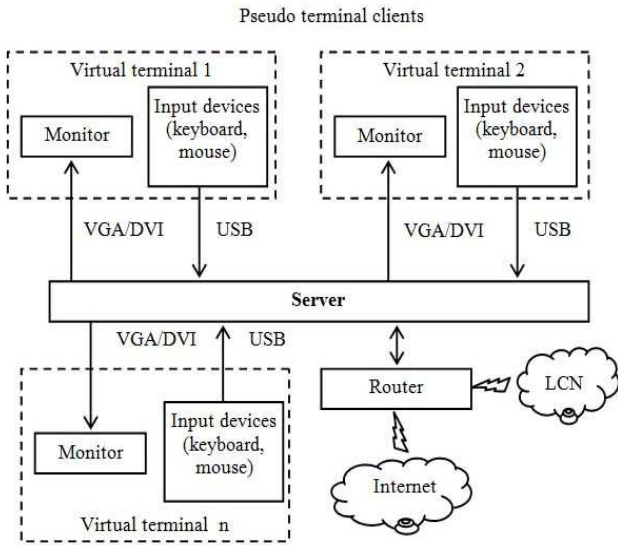


Figure 2. Pseudo terminal concept

When some user puts data in by means of a keyboard or other input devices connected with the input physical controller corresponding to this very user, virtual machines monitor does not capture data from the controller of data input, giving them to the processor of the user guest operating system.

Every operator works with the virtual machine via IO devices appointed to him/her and ideates that s/he is working on a separate dedicated computer.

Using the technology of virtualization it is possible to introduce mobile automated workstations based on tablet PCs, iPhon, etc. into the information system as well.

Software-hardware links are presented schematically in Fig. 3.

As such, the system provides “Desktop as a Service”, i.e. processing power of a computer with physically switched on hardware and working virtual machines monitor are “leased” to the remote user. The problem of multiple access with virtual machines is solved by appointing hardware to a virtual machine on demand of a user in real time mode as well.

#### IV. TECHNICAL MAINTENANCE

The main merit of the developed method for multi-user system organization is in the fact that it does not require modernizing of the existing computer arrangement architecture. The desired number of video adaptors and data input controllers are connected to the PCI-Express bus, while data IO devices necessary to each operator (monitors, touch panels, VR-headsets, keyboards, etc.) are set on the computer workstations and connected with the computer through transmission links of DVI interface for data output device and of USB interface for data input device (Fig. 4).

The number of terminals is limited by the number of video adaptors of the computational server; the number of terminals equals the total number of server video outputs. In case modern computational systems with 6 video adaptors are exploited, the number of terminals amounts 12.

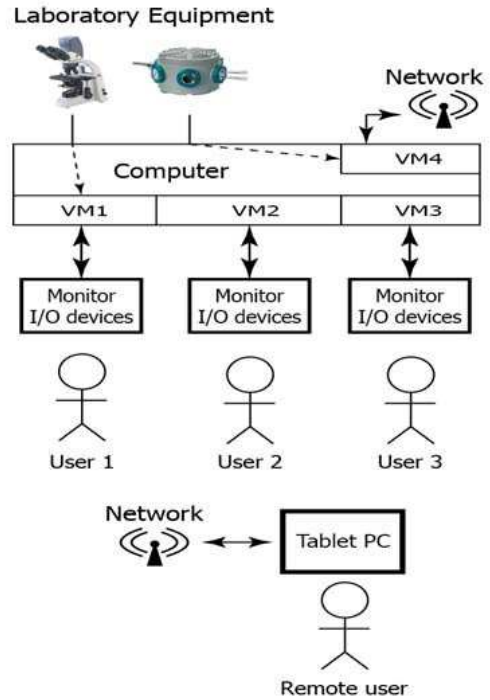


Figure 3. Software-hardware links

Each of the users works on an individual virtual machine served by a hypervisor. It should be noted that in case hypervisor Xen is used practically no limitations are imposed on the operational systems exploited by users. Moreover, they may be completely different depending on the requirements of the user.

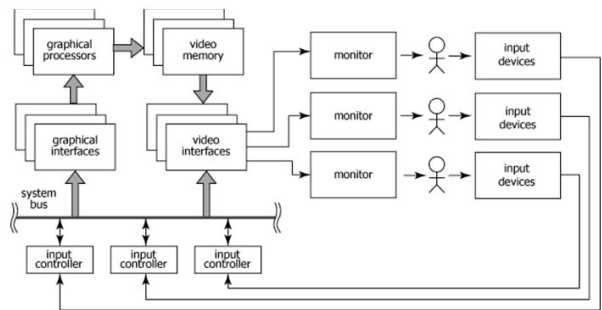


Figure 4. Virtualization technology for multi display systems

Linux remains the main software platform of the complex, however, it is only the host-system for the virtual machines monitor, i.e. it provides the work of the most part of the computer and hypervisor hardware.

#### V. CONCLUSION

Thus, the method of building multi-user multi-monitor systems can be used for creation of information educational systems for educational establishments of different levels.

Familiar alternative systems exploiting Useful MultiSeat Linux and Multiterminals technologies differ from the suggested one by fewer number of terminals (up to 10) and lower functional capabilities, which significantly limits the freedom of users in choosing software. In other words, alternative technologies, in contrast with the suggested one, make it possible that many users worked with one operating system simultaneously, which imposes limitations both on the software used and on the provision of the information security for the users.

The method elaborated provides convenient and efficient user work in the close to real time mode thanks to the lack of delays in the channels of data transfer between the automated workstations. It also allows to significantly lower the cost of the information system at the expense of more efficient resources usage.

Remote access broadens functional capabilities of the information system largely and improves the quality of distance education.

Besides, the suggested structure of building terminal systems with improved dynamic features can be used for technological processes control by means of a small number of users, for creating "islands of automatization", complex geographically-distributed automated control systems, as well as for creating computer workstations in computer-aided design systems.

#### ACKNOWLEDGMENT

The results presented in the paper are obtained with the grant of the Russian Foundation for Basic Research (RFBR) 10-08-97019-поволжье\_а «Развитие теории адаптивного интерфейсного согласования человека с технической системой» (Development of the theory of the adaptive interface coordination of a person with a technical system) (Petukhov I.) and with the grant support under the analytical departmental target program «Development of scientific of the higher school (2009-2010)» № 2.1.2/11701 "Investigation of the stability and ensuring the invariance of energy-efficient systems, adaptive control of dynamic objects" (Steshina L.).

#### REFERENCES

- [1] W.K. Giloi, J. Encarnacao, and S. Savitt, "Interactive graphics on intelligent terminals in a time-sharing environment," *Acta Informatica*, vol. 5, no. 4, pp. 257-271.
- [2] H.J. Lotz-Iwen and W. Steinborn, "The intelligent satellite-image information system ISIS," *Proc. AIP Conf.*, August 1993, vol. 283, pp. 727-734.
- [3] N. Mitrovic and E. Mena, "Adaptive User Interface for Mobile Devices," *Proc. of the 9th International Workshop on Interactive Systems. Design, Specification, and Verification (DSV-IS 2002)*, Springer Press, June 2002, pp. 47-61.
- [4] E. Edelhauser, "Client server versus distributed network applications in human resource management," *Proc. of the International Conference on Theory and Applications of Mathematics and Informatics (ICTAMI 2004)*, pp. 129-136.
- [5] C. Anderson and S. Greenberg, "Windows and UNIX on the Desktop: MetaFrame gives you the best of both worlds," *Windows & .NET Magazine*, May 2003, pp. 65-67.
- [6] J.E. Pezoa, M.M. Hayat, Zhuoyao Wang and S.Dhakal, "Optimal Task Reallocation in Heterogeneous Distributed Computing Systems with Age-Dependent Delay Statistics," *Parallel Proc. 39th International Conference (ICPP)*, Sept. 2010, pp. 111-120.
- [7] R. Green, *Expert CAD Management: The Complete Guide*. San Francisco: Sybex, 2007.
- [8] J. Halscott, *Desktop Virtualization and Evolving Strategies for IT Service Delivery*. San Francisco: Realtime publishers, 2011.
- [9] Y. Weimin, "Network turns green," *Huawei Communication*, Sep. 2009, vol. 51, pp. 36-38.
- [10] H. Ikebe, N. Yamashita, and R. Nishii, "Green Energy for Telecommunications," *Proc. INTELEC 2007*, pp. 750-755.
- [11] W.E. Crawford, "High Performance Controller Electronics for Networked Digital Displays," *Proc. SID conf. Americas Display Engineering and Applications (ADEAC 2006)*, SID Press, pp. 197-200.
- [12] I. Tanryverdiev, L. Steshina and I. Petukhov, "The hardware and software platform for automated control systems," *Proc. of the Junior Scientist Conference 2010*, April 2010, pp. 313-314.
- [13] I. Petukhov, L. Steshina and I. Tanryverdiev, "System of the distributed control and input of the information for automation of continuous technological processes in real time," *Bulletin of Mari State Technical University*, vol. 1, 2009, pp. 72-80.

# A Generic Approach to Interactive University Timetabling

Michael Zeising, Stefan Jablonski  
 Chair for Applied Computer Science IV  
 University of Bayreuth  
 Bayreuth, Germany  
 {michael.zeising, stefan.jablonski}@uni-bayreuth.de

**Abstract**—Room utilization is a problem for universities most notably when students do not have to enroll for the courses they want to attend. In this case, the only information available is the examination regulation of their respective course of study. In this paper, it will be illustrated how student constraints can be inferred from these regulations and how they can be used to reduce the task to a well understood problem. Furthermore, an innovative user interface based on these constraints which enables highly interactive university course scheduling will be presented. In order to support a wide range of environments, the approach rests upon a very general domain model and does not depend on a specific solver technology.

*Keywords*—interactive timetabling, decision support systems, examination regulations, planning user interfaces

## I. INTRODUCTION

Universities are forced to optimize their room utilization while satisfying time constraints for students as well as for teachers. The task becomes even more complex if the university decides to not force their students to enroll for courses before attending them. The only information from which time constraints for the students may then be derived is the curriculum to which they committed themselves. The actual scheduling problem is well understood and numerous mathematical approaches have been published so far [1-3]. In the scope of these approaches, a curriculum is understood as set of courses that a student must attend [4]. However, for German universities this abstraction is oversimplified. There are many paths through a course of study and only a few courses could be identified as a mandatory core. The possible paths that lead to a degree are constrained by often complex regulations.

This paper demonstrates how curriculum-based scheduling can be significantly improved by incorporating these regulations into the planning problem. Furthermore, it will be shown how an innovative user interface allows for an interactive scheduling. By the use of simple technology-agnostic models, the approach imposes only very few restrictions on the surrounding landscape and can therefore be easily integrated into existing management solutions. The suggested approach is also not bound to a specific optimization technique.

## II. PROBLEM DESCRIPTION AND REQUIREMENTS

A university maintains hundreds to thousands of rooms shared for teaching activities. To make the best possible use of these rooms, they are managed in a centralist manner by at least one room authority. A room authority is responsible for positioning a teaching activity in time and space so that no hard constraints and as few as possible soft constraints are violated. An authority, however, is not permitted to alter an activity's properties like its duration. It is argued that this planning procedure must be of interactive nature. So the challenge consists in providing the best possible support for human planners at the user interface level.

The actual constraints may vary between universities and therefore must be specified individually. It is important for the solution to be independent of the respectively used optimization technology. So the second challenge is to make sure that the solution exposes a generic and yet efficient interface to solver engines and allows for as many kinds of constraints as possible to be implemented.

The most important hard constraints usually encompass persons and rooms being impartible. While the violation of most of them may be easily discovered, there is one very complex constraint: time overlaps for students. In this context, a university offers courses of study that lead to certain degrees. The structures and rules of such a course of study are specified by regulations (also called examination regulations). For easier naming, it is assumed that *one* regulation describing one course of study leading to one degree for a student. Regulations are usually subject to change so they are versioned but it is expected that one student is always registered for only one regulation (version) at a time. The third task is to exploit the information on students for inferring potential timetabling conflicts.

## III. CONTRIBUTION

Within this paper, insight into how university course timetabling can be improved by interpreting basic regulation rules will be delivered. A generic approach to an interactive regulation-based planning tool will be introduced. By the term “generic” it is meant that the models used are simple and generally applicable to universities. “Interactive” means that the goal is a decision support system for planning authorities. By means of an innovative user interface the solution will support the authorities with the intelligence of a



semi-automatic optimization. The tool presented will not depend on a specific optimization technology.

In Section IV, the domain model of the application for adequately capturing regulations, the teaching offer and planning information will be introduced. In Section V, it will be demonstrated how these information can be used to infer conflicts for the students. In Section VI, a model used for communicating with the solver for automatic optimization will be proposed. Finally, in Section VII, the planning process supported by the approach and a user interface for interactive planning will be presented.

#### IV. DOMAIN MODEL

The essential information on which the approach is based have to be adequately modeled. In the following, a brief overview of this data model will be given.

##### A. Regulations and Teaching Offer

The first task consists in modeling regulations. The authors have gained the experience that they may be modeled as trees of teaching units where the leaves are courses and the inner nodes may declare constraints on their children. The left side of Figure 1 shows that a regulation consists of units, which can be “phases”, “areas”, “sections”, “stages” and so on. Everything that is used to structure courses of study is considered a unit. The module is a special kind of unit which holds a number of credit points gained according to the European Credit Transfer System (ECTS) if the module is passed [5]. A module may contain courses, which again are special units. The course holds an amount of workload in hours per week which is interpreted as the time of attendance for students and teachers. It is assumed that a course always lasts for one semester which, up to the knowledge of the authors, holds for every German university. If a course spanned more than one semester it

would be split up into several continuously numbered parts like “Basic Mathematics I” and “II” and offered in different semesters.

The model allows for two very simple rules: prerequisites for a unit and choices. Each unit may reference others that must be attended before the unit because e.g., it requires knowledge gained in the other units. Generally, the hierarchy of units is interpreted as connected neither by “and” nor by “or” operators. That means that for a unit containing two or more others *some* of these child units are expected to be attended. For modeling explicit choices instead, the `nOutOf`-attribute may be set, which means that exactly *n* out of all child units need to be attended by a student.

It is assumed that every student is registered for exactly one regulation. Students registered for multiple regulations (and therefore enrolled in multiple courses of study) at the same time do not raise a problem but they cannot be supported. However, this should be an acceptable restriction as studying multiple independent courses of study is a rare case.

The actual teaching offer per semester is modeled by the right half of Figure 1. It states that a course mentioned in a regulation may be offered in a semester. A course offer in turn consists of at least one session which involves at least one person usually in the role of a teacher.

This model is, at the best of the authors’ knowledge and experience, in line with the respective models of major commercial university management solutions. Therefore, the approach can be simply integrated with these systems.

##### B. Time, Space and Planning

There is only one valid planning at a time. A planning consists of allocations while an allocation assigns a session to a certain room and a certain time. The solution is not bound to a specific optimization model and time granularity.

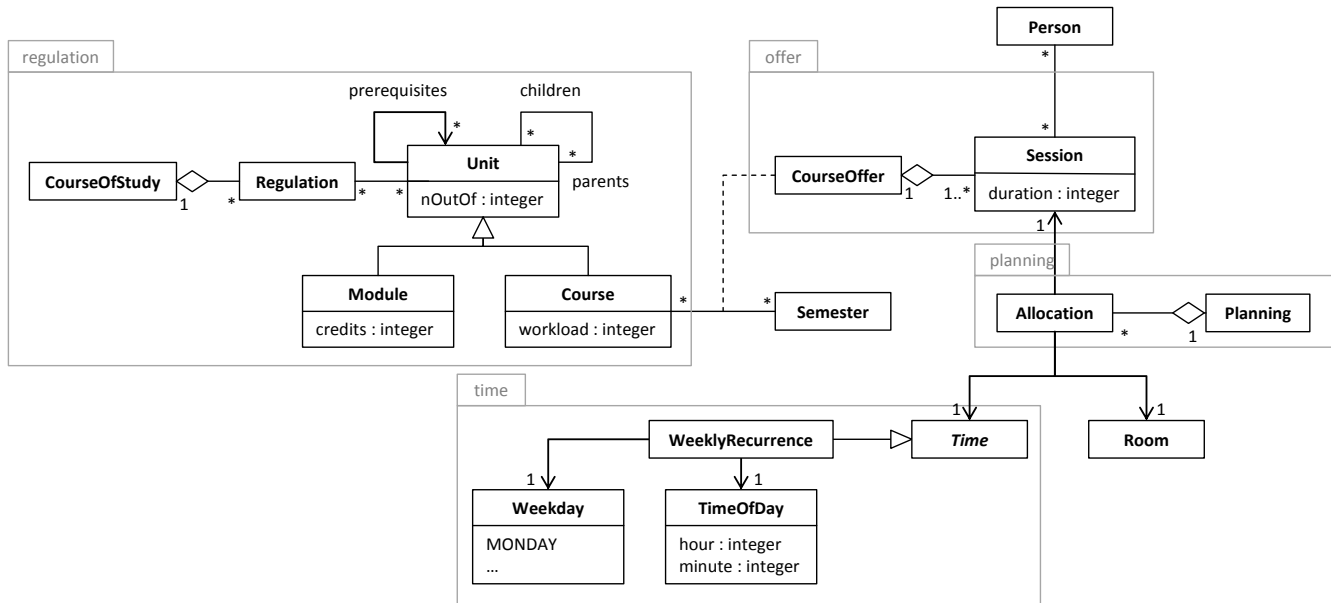


Figure 1. Domain model for capturing regulations, teaching offers and plannings

As a result, time is represented absolutely and technology-agnostic within the data model. Time may be a single event (not shown) or a recurrence rule. In the most frequent case, sessions will occur weekly. Simple weekly recurrence rules may be modeled as shown in the lower half of Figure 1. There is an event with a certain duration recurring every certain weekday at a certain time of day. Of course, these rules quickly become complex as an event may only occur every second or third week and there may also be gaps within the rule but we concentrate on the simple case for now.

V. DERIVING SESSION CONFLICTS

A simple perception of collisions and conflicts is used: two sessions “conflict” if they must not overlap in time. If two sessions conflict and overlap, they “collide” and render a timetable either unfeasible or worse than another.

There are the usual trivial cases of two sessions conflicting that most approaches support and that may easily be determined. Therefor two sessions are conflicting if

- they take place in the same room or
- they involve the same teacher.

Determining whether two sessions are conflicting from the students’ point of view, however, is nontrivial and, up to the authors’ knowledge, not supported by any existing solution (if not provided to the system *a priori*). From the students’ point of view, two sessions are potentially conflicting if

- they are offered in the same semester and
- there is a regulation containing both of them and
- there is at least one “path” allowing for attending both of them.

In short, they conflict if there is potentially at least one student attending both of them while following his or her regulation.

For illustrating the principles of deriving conflicts from a regulation, imagine two exemplary regulations shown in Figure 2.

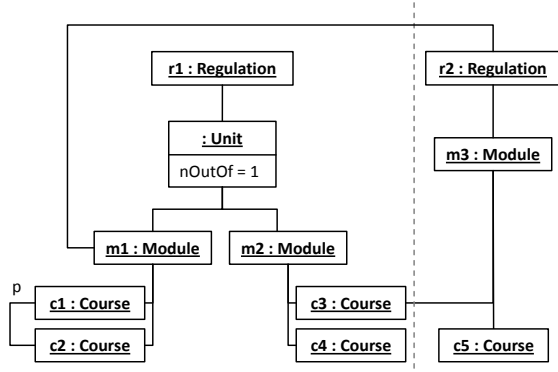


Figure2. Object model for two exemplary regulations

Regulation  $r_1$  consists of a section containing two modules from which one has to be chosen. Each of the modules contains two courses. Course  $c_2$  has  $c_1$  as a prerequisite which means that  $c_1$  has to be attended before  $c_2$ . Regulation  $r_2$  reuses module  $m_1$  and contains another module

$m_3$ . Module  $m_3$  reuses course  $c_3$  and contains a further course  $c_5$ . This degree of reuse between regulations is typical for modular interdisciplinary courses of study at the university. Figure 2 shows the resulting object graph for the example in the form of an object diagram.

A. Semester-specific Representation of Regulations

Now, assume that during the considered semester there is one session  $s_i$  offered for each course  $c_i$ . For being able to operate directly on sessions, the object graph depicted in Figure 2 is reduced to a specialized representation of the regulations shown in Figure 3. This representation may be efficiently computed when using adequate representations of hierarchical data like *nested sets* [6] on the storage level.

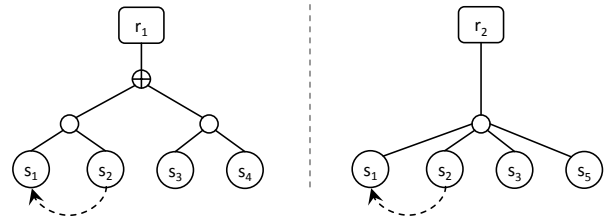


Figure 3. Alternative representation of the exemplary regulations

This form reveals the interpretation of the regulations. For regulation  $r_1$  the two paths are connected by an *exclusive or* while for  $r_2$  all sessions are just grouped. The empty node is interpreted neither as *or* nor as *and* because this part of the regulation is considered “under-specified” meaning that potentially all of the sessions are attended by the same student.

B. Local Conflict Handling

Initially, it is assumed that all sessions within a regulation are conflicting. This results in the two conflict graphs in Figure 4a for the two regulations.

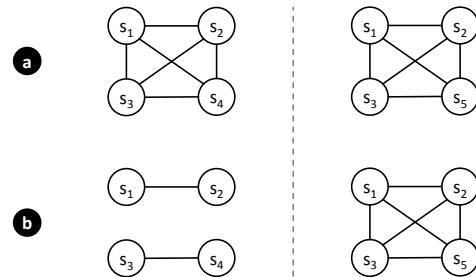


Figure 4. Local conflict graphs resulting from the regulations

Other student-aware approaches would be finished at this point. But now the *exclusive or*-connector between the two session groups in regulation  $r_1$  is interpreted. It can be safely assumed that if a student has to choose between the two directions there will never be a student (registered for this regulation) attending both of them. So the according conflict edges may be removed which leads to the graphs shown in Figure 4b. At this point, it becomes clear why this step has to be performed regulation-locally. There could be another regulation where the modules or courses are reused and no exclusive choice has to be made between them. Actually,

regulation  $r_2$  also contains  $s_1, s_2$  and  $s_3$  but without a choice so the conflicts must not be removed there. The meaning of the two graphs in Figure 4b now reads as follows:

- there may be at least one student enrolled in  $r_1$  who attends  $s_1$  and  $s_2$  in this semester and at least one who attends  $s_3$  and  $s_4$  and
- there may be at least one student enrolled in  $r_2$  who attends  $s_1, s_2, s_3$  and  $s_5$  in the respective semester.

C. Global Conflict Handling

The individual conflict graphs of the regulations can now be merged to one graph for further steps. The set of vertices of the global graph is the union of the local sets. It contains an edge if any of the local graphs contains it. The result for our example is shown in Figure 5a. Note that some of the edges removed due to the *exclusive or* are restored after the merge.

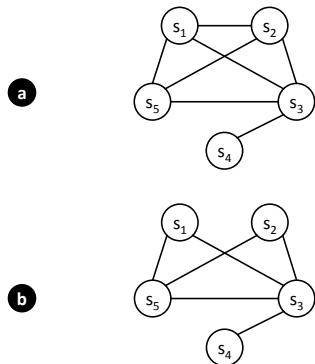


Figure 5. Global conflict graph resulting from the regulations

Further edges may be removed by interpreting the prerequisites of a unit. Figure 1 shows that a unit may declare other units as its prerequisites which means that these units must be attended by a student before the unit. This is a global relation which does not depend on regulations and may not be altered by them. As it is assumed that units span at least one semester, it may be inferred that if a unit is a prerequisite of another one then there will never be a student attending both of them in the same semester and there is no conflict between them. In the example, course  $c_1$  is a prerequisite for  $c_2$ . This has been projected to  $s_1$  and  $s_2$  so the conflict between these two sessions is removed leading to the final conflict graph shown in Figure 5b.

Please note that by interpreting only very simple properties of the regulations, the number of conflicts has been reduced by almost 40%. Due to fewer conflicts, the optimization procedure now has more degrees of freedom and a better planning becomes possible.

VI. INTEGRATING AN OPTIMIZATION ENGINE

After the potentially conflicting sessions have been identified, the task can be reduced to a general optimization problem. This general problem known as *university course timetabling* may now be solved by one of the numerous approaches mentioned before [7]. For the solution to be independent of the chosen technology, the information necessary for optimization is translated to a planning model.

The model constitutes a (not necessarily feasible) solution of the problem and contains information needed for implementing constraints like, e.g., [4]:

- *impartible teachers*: a person may only be involved into one session at a time
- *impartible rooms*: there must be only one session taking place in a room at a time
- *room capacity*: the expected audience of a session must not exceed a room’s seating capacity
- *timetable compactness*: there should be no time gaps within a planning
- *room stability*: when assigning a session to another room, the distance to the original one should be minimal
- *session conflicts*: conflicting sessions must not overlap in time

Figure 6 shows the core of the planning model. The underlined properties of an allocation (room and slot) are to be changed by the solver during optimization.

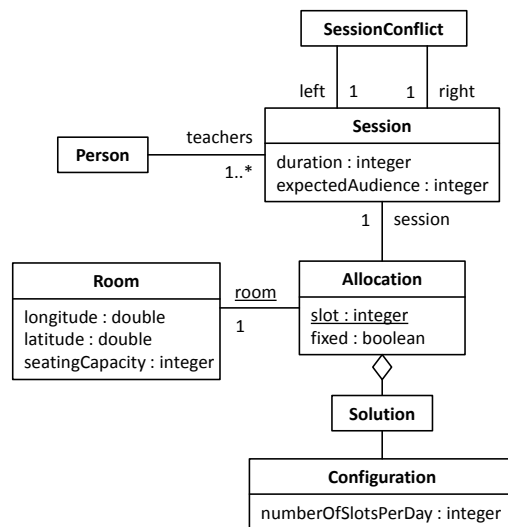


Figure 6. Planning model for communicating with a solver

For the current prototype, a solver implemented on the basis of a linear programming method is employed but it is important to note that the solution is not bound to any particular optimization algorithm or engine. The solver backend is completely exchangeable by design. This is a crucial point because high performance solvers for large real-world problems and the accompanying hardware usually impose significant costs. With an independent solution the decision for an appropriate licensing model can be made by the university individually and even custom developed solutions may be easily integrated.

VII. SUPPORTING INTERACTIVE PLANNING

It has been illustrated how to infer conflicts for students between teaching sessions. In the following, it will be shown how these information serve as the basis for a “student dimension” and enable a comprehensive planning user interface.

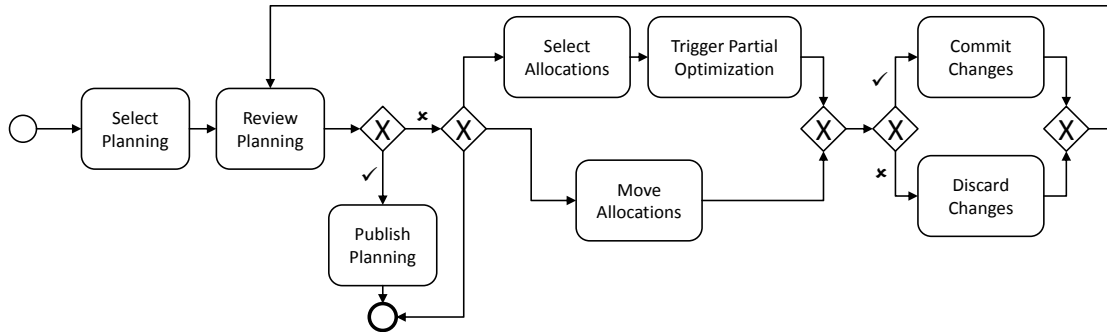


Figure 7. Planning process supported by the user interface

Course scheduling is a highly political topic not only in Germany [8] involving a lot more aspects than the ones currently modeled in university management software [9]. In [3] the authors actually complain that “practical course timetabling is 10% graph theory, and 90% politics” [3]. Therefore, we argue that course scheduling must be a semi-automatic process that can only be successful when combining the users’ knowledge with a decision support system. Figure 7 shows a BPMN representation of the planning process skeleton as our interface supports it [10].

First of all, a planning is selected for editing. This may be an already valid and published planning or an intermediate sketch. The step that surely makes the highest demands on the user interface is the review of the edited planning. The human planning authority needs to capture information on all the eventual collisions like:

- a room is occupied multiple times
- a room’s capacity is exceeded
- sessions involving the same teacher(s) overlap
- sessions potentially attended by the same students overlap

In order to support these aspects, we have designed an innovative resource-time-view. It arranges session allocations on a resource-time-grid. Allocations are colored

according to whether they do not, potentially or certainly collide with others. Colliding sets can be quickly identified by hovering over them. The view is implemented as a JavaServer Faces component so that it can be used within a Java EE environment, a standard platform for business applications. In addition to that, it is fully amenable to the top three recent browser technologies. Figure 8 shows the detail of a screenshot taken from the view in action.

If the planning turns out to be feasible it may be published and made the valid planning, if not the process may just be aborted and deferred or the semi-automatic editing phase may be entered. For manual editing, the planner may move allocations directly in time and space per drag and drop. Of course, this manual planning is only intended for small easily manageable changes of the schedule. If greater numbers of sessions should be moved the planner may resort to the automatic planning of partial timetables. For this to accomplish, he/she selects the sessions to be moved by clicking on them and triggers the optimization. The sessions are moved and the resulting difference is visualized by arrows and can be applied to the edited planning afterwards.

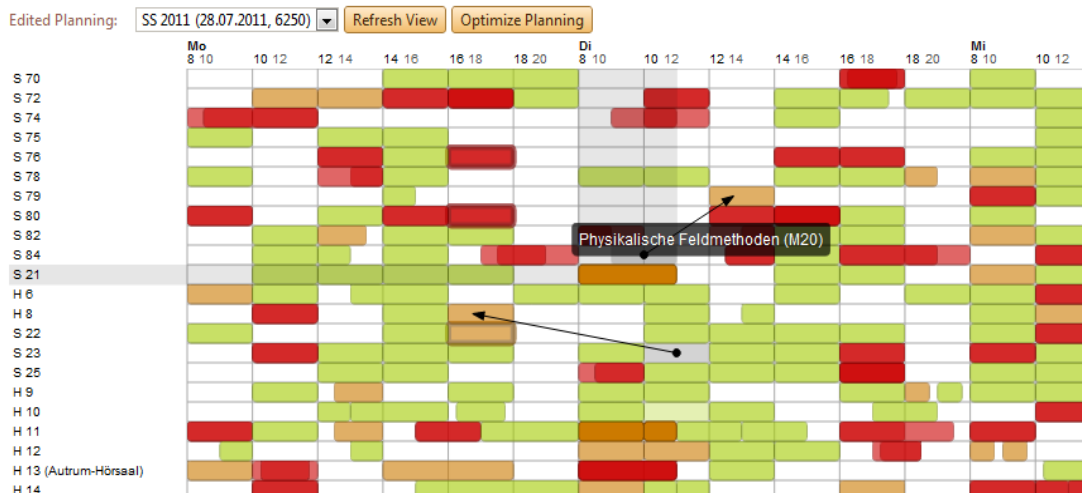


Figure 8. Screenshot of the interactive planning view

## VIII. EVALUATION

A prototype of the solution has been integrated with the system landscape of the University of Bayreuth. Regulations are provided by the *FlexNow* examination management system [11]. Organizational and facility information as well as course information are provided by the *HIS LSF* system [12]. The prototype is currently being evaluated by the central room planning bureau of the university. The next step consists in making the results available to other universities.

## IX. RELATED WORK

Most research effort is spent on the autonomous optimization of generally acknowledged “standard” problems like, e.g., *post enrolment based course timetabling*, *curriculum based course timetabling* or *examination timetabling* [4]. The assumption is that real world problems of universities may all be reduced to one of these models. However, to the best of the authors’ knowledge, there is no approach that examines whether the optimization can be simplified or improved if further knowledge like regulations is provided.

There are only very few approaches to interactive university timetabling. In [13], Piechowiak and Kolski present an interactive system supporting a resource-time-view adaptable to the needs of different kinds of users (namely the “designer”, the “analyzer” and the “consultor”). Like others [14], the approach is based on a model that relates students by fixed groups. This grouping must be made available to the system *a priori* in the form of syllabi which are usually not available in universities. Besides this, the system is bound to a specific solver technology.

Though the solution presented in [15] supports interactive timetabling, it is based on an inappropriate school model of students arranged in classes. Moreover, it does not provide a comprehensive view on time and the resources involved. It is also bound to a specific solver implementation.

## X. CONCLUSION AND FUTURE WORK

In this paper, it has been demonstrated how information gained from examination regulations facilitate comprehensive interactive university course timetabling. It has been explained how to model these regulations in a way so that the information can be gathered from adjacent university management solutions. By the use of a lifelike example, it has been showed how to prepare the information for a generally acknowledged and well examined timetabling problem. The obtained constraints were used to integrate an optimization engine in a technology-agnostic way and to enable a highly productive user interface supporting human planners.

The authors are currently investigating how further rules within regulations can be employed for reducing student-related conflicts. They also concentrate on the weighting of these conflicts based on the expected number of students affected. Furthermore, the solution is currently extended on the planning of non-weekly events like block courses and conferences held at the university. These events must be

planned on the basis of the weekly teaching events because they must not collide with it. In parallel to these aspects, the aim is to evaluate the approach at other universities.

## ACKNOWLEDGMENT

This work is kindly funded by the Oberfrankenstiftung grant for project number 02803.

## REFERENCES

- [1] O. Rossi-Doria, M. Sampels, M. Birattari, M. Chiarandini, M. Dorigo, L. M. Gambardella, "A Comparison of the Performance of Different Metaheuristics on the Timetabling Problem," in 4th International Conference on the Practice and Theory of Automated Timetabling (PATAT 2002), 2003, pp. 329-351.
- [2] S. Abdullah, E. Burke, and B. McCollum, "An Investigation of Variable Neighbourhood Search for University Course Timetabling," in 2nd Multidisciplinary International Conference on Scheduling: Theory and Applications (MISTA), 2005.
- [3] M. W. Carter, "A Comprehensive Course Timetabling and Student Scheduling System at the University of Waterloo," in 3rd International Conference on the Practice and Theory of Automated Timetabling (PATAT 2000), 2001, pp. 64-84.
- [4] B. McCollum, P. McMullan, B. Paechter, R. Lewis, A. Schaerf, L. Di Gaspero, "Setting the Research Agenda in Automated Timetabling: The Second International Timetabling Competition," *INFORMS Journal on Computing*, vol. 22, 2010.
- [5] E. Commission, *ECTS Users' Guide*. Luxembourg: Office for Official Publications of the European Communities, 2009.
- [6] M. J. Kamfonas, "Recursive Hierarchies: The Relational Taboo," *The Relational Journal*, 1992.
- [7] D. de Werra, "An introduction to timetabling," *European Journal of Operational Research*, vol. 19, pp. 151-162, 1985.
- [8] B. McCollum, "The implementation of a central timetabling system in a large British civic University," in *Practice and Theory of Automated Timetabling*, 1998, pp. 237-253.
- [9] A. Bode and R. Borgeest, *Informationsmanagement in Hochschulen*: Springer Berlin Heidelberg, 2010.
- [10] I. Object Management Group, "Business Process Model and Notation (BPMN) Version 2.0," ed, 2011.
- [11] W. I. f. H. d. U. Bamberg. (12.08.2011). *FlexNow*. Available: <http://flexnow.uni-bamberg.de/>
- [12] H.-I.-S. GmbH. (12.08.2011). *Lehre, Studium, Forschung*. Available: <http://www.his.de/abt1/ab10>
- [13] S. Piechowiak and C. Kolski, "Towards a generic object oriented decision support system for university timetabling: an interactive approach," *International Journal of Information Technology and Decision Making*, vol. 3, pp. 179-208, 2004.
- [14] J. P. Pedroso, N. Moreira, and R. Reis, "A Web-Based System for Multi-Agent Interactive Timetabling," in *International Conference on Knowledge Engineering and Decision Support (ICKEDS'04)*, 2004, pp. 1-6.
- [15] T. Müller and R. Barták, "Interactive Timetabling: Concepts, Techniques, and Practical Results," in 4th International Conference on the Practice And Theory of Automated Timetabling (PATAT 2002), 2002, pp. 58-72.

# The Utility of Controlled Vocabularies within Bookmark Management Tasks

Siu-Tsen Shen

Department of Multimedia Design  
National Formosa University  
Hu-Wei 63208, Taiwan, R.O.C.  
[stshen@nfu.edu.tw](mailto:stshen@nfu.edu.tw)

Stephen D. Prior

School of Engineering and Information Sciences  
Middlesex University  
London N14 4YZ, UK  
[s.prior@mdx.ac.uk](mailto:s.prior@mdx.ac.uk)

**Abstract**—This research investigates the utility of adopting a controlled vocabulary approach to bookmark management. An initial user survey conducted for this research has shown that just over half the population use bookmarks to save important websites and that 75% of these people use up to three sub-levels only. The bookmark facility within all current web browsers is therefore underutilized and the argument that users need and want greater freedom and flexibility to create their own unique file structure is disputed. We conclude that users need a simple, logical and contextual system of bookmark management which complements their daily lives.

**Keywords** - *controlled vocabulary; bookmark management; web browser; information search; information retrieval.*

## I. INTRODUCTION

Personal file management has become significantly more important as the daily amount of digital data we view and store on our computers, smart mobile devices and web browsers increases. The familiar concept of the hierarchical file system allows us to group our important information in an organized tree structure, i.e., the use of folders, referred to as directories within folders. Folders (directories) normally include other possible sets of files or sub-folders (subdirectories). This is useful for desktop organization and bookmark management in relation to relevant topics and information.

Nevertheless, the usefulness and efficiency of the hierarchical file system has been debated over the last few decades [1-3]. Bloehdorn and Volkel (2006) stated that current file systems are problematic (with their single location ascribe) to browse to maximum specificity (retrieval needs the exact directory), miss-orthogonality (all orthogonal dimensions being forced into one access path), path order dependence (contrast to the directories seen as independent attributes), have no query refinement (no list of relevant directories to search), and have no navigational aid (no indication of the content of subfolders) [4].

In Section II, we will discuss the related work with particular references to controlled vocabularies. Following this section will be data on a preliminary user study. A discussion of the initial findings, analysis, and conclusions follows on from this.

## II. RELATED WORK

Tags, also known as metadata, have been widely used within Web 2.0, and serve as labels to be easily-identified in information retrieval tasks. Successful examples include del.icio.us [5], the social bookmarking website, Flickr [6] for image collection and YouTube for video collection. Compared to the inflexible, one-way system of the hierarchical file system, a tagging system gives the user a great deal of freedom to mark their wanted items, which could be multiply-tagged. Especially popular, is the social and collaborative sharing of information, also known as Folksonomy [7]. Websites with a large collection of text listed tags in alphabetical order (normally) shown as tag clouds are helpful for browsing and searching by their various fonts, size and colour. A successful social tagging website like ‘43 Things’ is an example. Nevertheless, users may create thousands of tagged items and end up spending more effort on sorting and finding the tags that are needed.

### A. Controlled Vocabularies Applied to IT

In contrast to the concept of Folksonomy, which is an informal and liberal collaborative tagging system, a controlled vocabulary is a restricted system of textual tags normally used for large datasets. Examples include the Library of Congress Subject Headings (LCSH) [8], the European Patent database [9] and the Yellow Page phonebook. For example, the International Patent Classification (IPC), established by the Strasbourg Agreement 1971 [10], provides for a hierarchical system of language independent symbols for the classification of patents and utility models, according to the different areas of technology to which they pertain. In total, this consists of eight main subject headings under which every patent application has to be categorized [11]. These are listed below:

- Section A — Human Necessities
- Section B — Performing Operations, etc
- Section C — Chemistry; Metallurgy
- Section D — Textiles; Paper
- Section E — Fixed Constructions
- Section F — Mechanical Engineering, etc

Section G — Physics  
 Section H — Electricity

The purpose of controlled vocabularies is to classify the terms such as words or phrases defined by experts or authorities in order to make retrieval performance more efficient. However, it requires a certain level of preciseness in the interpretation of the terms. It is common that users experience a familiar situation, whereby they conduct an online search by typing in keywords which may have a more general and broad meaning, and might therefore come up with a long list of irrelevant or unwanted information. A successful performance when using a search engine normally relies on the individual user’s capability as to whether he or she could select the appropriate keywords or not. Controlled vocabularies are thus generally applied to thesauri, taxonomies and ontology [12].

Several researchers concluded that people’s capability of categorizing information is cognitively difficult [13-16]. It has been stated that human’s ability to categorize is hard to identify and is definitely not in a strict hierarchical structure, but shall be assumed to be more fluid and flexible [17, 18]. File systems, rigorous hierarchical mechanisms, such as ‘My Documents’ and ‘My Favorites/Bookmarks’ have been shown to have the usability problems of usefulness and appropriateness, including filing management, document organization, and document retrieval [2, 19-22].

Several attempts to use different approaches to replace the standardized hierarchical system have provided solid results in previous research. Barreau and Nardi (1995) found that people prefer to use location-based search and visual grouping, rather than complex data structures [1]. Gifford, Jouvelot, Sheldon, and O’Toole (1991) employed an associative attribute-based approach to access files and directories by semantic indexing, and proved to be more effective than the hierarchical structure [23]. Dourish et al. (2000) implemented a property-based approach to amend the traditional file system’s problem with a uniform framework [2].

III. PRELIMINARY USER STUDY

In order to gain an understanding of user behavior, with regards to the use of current file systems and browsing patterns, a preliminary study was conducted online via an academic-based social blog in August 2011. The recruitment of participants aimed for experienced computer users, in that they could provide more insightful views according to their intensive usage on task performances such as searching, browsing and organizing information. There were a total of 60 participants, consisting of 37 females (62%) and 23 males (38%). Their age was between 18 to 25 years old, with 50% of them being aged 20 years old (see Figure 1).

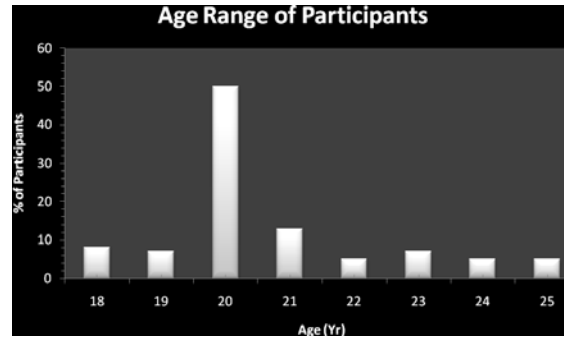


Figure 1. Age range of participants.

The group consisted of 88% of the participants who were University students, 10% from graduate schools, and 2% with a senior high school degree. In terms of computer experience, 33% of the participants had 10-12 years of experience, 22% with 8-10 years, 17% with 14-16 years of experience, and 12% with 12-14 years of experience (see Figure2).

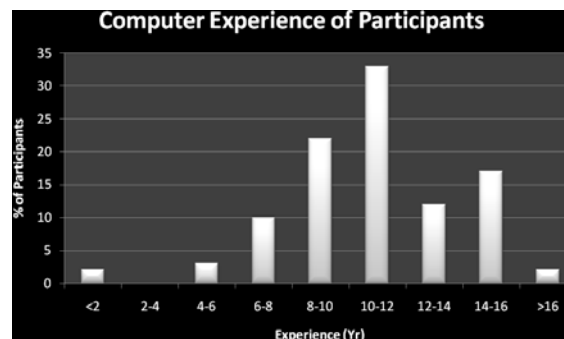


Figure 2. Computer experience of participants.

In terms of their Internet usage, 28% of the participants spent 6-8 hours, a further 23% of the participants spent 4-6 hours and 8-10 hours, and 12% participants who spent more than 12 hours a day online (see Figure 3).

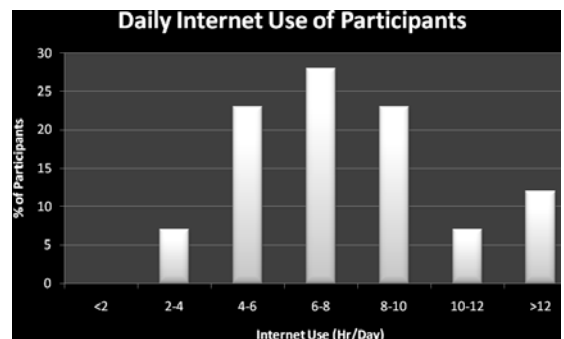


Figure 3. Daily internet use of participants.

The majority of the participants (93%) did search for information from their Bookmark folders, whilst only 7% of the participants did not. There were 83% of Bookmark

folders who had less than 50, 13% who had 50-100 folders, and 3% who had 100-150 folders (Figure 4).

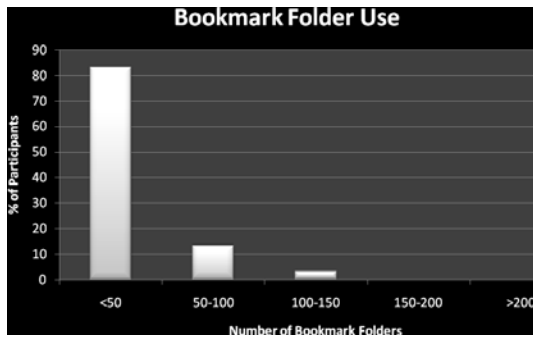


Figure 4. Bookmark folder use.

With regards to the type of computers used, there were 83% PC users, 7% Mac users, and 10% users who used both PC and MAC. 73% of participants use Google Chrome as their default browser, 15% with IE, and 12% with Firefox. Most of the participants had installed the latest versions of their browsers: 60% with Google Chrome 13, 12% with Firefox 5, 10% with IE 8 and 9. Nevertheless, there were 18% of the participants who did not know the version of their browser.

In terms of the use of Bookmark sub-folders, it was found that 53% of the participants did use them, and 47% who did not use them at all.

With regards to the maintenance of folder levels, 20% of the participants used 3 levels, 13% used 2 levels, 10% used 4 levels, 7% used only one level, and 3% used 5 levels (see Figure 5).

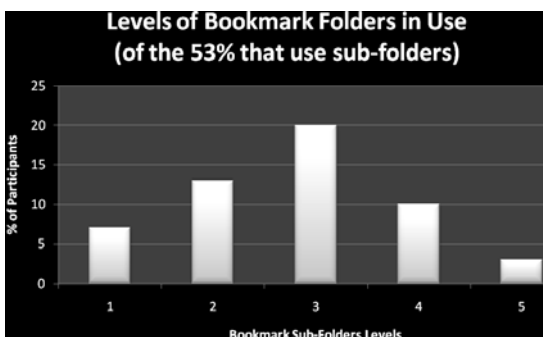


Figure 5. Levels of bookmark sub-folders in use.

#### IV. DISCUSSION OF INITIAL FINDINGS

From the outset, it was a desirable goal to have even numbers of male and female participants. However, the study was conducted online and therefore the researchers could not foresee or manipulate the numbers from the participants, since it was freely available 24-7. The vast majority of the participants were university students. Although 10% of the participants were from graduate schools, they were studying related digital media subjects. It

is important to note that they are dependent on computers, in that computers are the essential tools for their future career prospects. It is surprising to note that even in a group of young people such as these; one third of the participants had 10-12 years of computer experience which corresponds to the prevalence of Internet boom back in late nineties. It was found that about 75% of the participants spent between 4-10 hours/day using the Internet, which indicates the strong need for information and communication tools. It is not surprising that the PC is still dominating the market. However, with the latest popularity of multi touch devices like the iPad, it is still hard to tell when these will filter down to common use. With the recent trend for using smart phones, it was not surprising that 30% of our survey used one; there were 20% of the participants who used the HTC android platform, compared to 7% for iPhone users and 3% for Blackberry users.

A surprisingly high number of participants (70%), used Google Chrome as their default browser (even the Mac users); this indicates the advantage for web browser which are integrated within powerful search engines. Most of the participants kept up-to-date with the latest versions of their browsers, which also reflect their fast adaptation to the newest technology available. The frequency of using file systems under Bookmarks to search for information and files proved that the vast majority of the participants were indeed making efforts to organize their data. However, a majority of the participants managed their files into less than 50 folders.

It is interesting to note that 47% of the participants did not use sub-folders to organize their Bookmark files into more refined levels even though they created main folders to store information. Furthermore, 75% of the participants who used sub-folders managed to organize their personal information using 3 levels or less. This suggests that the current file systems that give almost unlimited creation of folders and sub-folders are perhaps unnecessary, because they are not well utilized by users who may either not want to make much effort on sorting their database, or may be aware that they might not be able to retrieve their desired information efficiently. Even high tech users find it difficult to manage the overwhelming data tsunami which hits us all on a daily basis; resorting to either not bookmarking or bookmarking without using folders and sub-folders.

#### A. Limitations of the study

This study can be criticized for only having a small number of homogenous participants and an uneven balance of genders represented. However, it is important to note that all the participants were studying multimedia design relevant subjects which require professional skills within several advanced software, as the aim of this study was focused on the experienced computer user.

Furthermore, this study was conducted via an online process and offered no cash in return for participation; therefore it was hard to manipulate the exact even numbers of both males and females compared to a lab-controlled setting environment. It is interesting to note that the majority



of participants were aware of the traditional hierarchical file system and did use them. However, nearly half of the participants did not create any sub-folders. Either they did not have a habit of organizing information or they thought that the traditional file system structure might not be helpful.

Traditional file system structures (developed by IT specialists) have existed for several decades and provide unlimited freedom for creating files and folders for users to organize their personal information. Based on our findings, three quarter of the participants used files and directories only up to 3 levels. This suggests that the current bookmarking system with 255 levels may be overly complex levels. This could be made simpler and more intuitive in terms of categorization via controlled vocabularies.

The purpose of getting online is mainly for searching, socializing and communication. Users may not appreciate the hierarchical filing system as others do. Moreover, it requires a lot of time and effort in organization information and does not guarantee users could successfully retrieve their required data when needed. Therefore, if we take this notion further, it would be a better idea to find an adequate approach to make the existing file systems into a simplified and deductive knowledge repository.

#### V. CONCLUSION

Based on our literature review, together with the results of the preliminary study listed above, several usability problems have been identified. The elements in need of improvement are: categorization, optimum levels of sub-folders, ambiguity of the use of vocabulary and contextual user mental models.

It is anticipated that if we could make the filing system less complicated and less strict, it would encourage users to be more willing to organize their information under such architecture. It is not our intention to replace the existing file system, but rather to offer a fresh perspective in visualization and user interface design.

The use of controlled vocabularies to assist in structured information storage and retrieval tasks looks to be promising, yet due to the natural ambiguity of descriptions of any specific term or object, it may appear not sufficient enough to achieve adequate understanding of precise meaning.

From an analysis of people's daily lives, further work is proposed to use a controlled vocabulary, which is divided into four primary facets, i.e. Work, Home, Travel, and Health. These could each further be broken down into a secondary level of say ten sub-categories. The use of these primary and secondary facets could help users reduce confusion and simplify the procedure when they organize their web information using bookmarks.

#### REFERENCES

1. Barreau, D. and B. Nardi (1995) *Finding and Remembering File Organization from the Desktop* SIGCHI Bulletin. **27**(3): pp. 39-43.
2. Dourish, P., W. Edward, A. LaMarca, J. Lamping, K. Petersen, M. Salisbury, D. Terry, and J. Thornton, *Extending Document Management Systems with User-Specific Active Properties*. ACM Transactions on Information Systems, 2000. **18**(2): pp. 140-170.
3. Sullivan, K. *The Windows 95 User Interface: A Case Study in Usability Engineering*. Proceedings SIGCHI. 1996: ACM: pp. 473-480.
4. Bloehdorn, S. and M. Völkel. *TagFS - Tag Semantics for Hierarchical File Systems*. WWW2006. 2006. Edinburgh, UK.
5. AVOS. *Delicious*. 2011 [retrieved: December, 2011]; Available from: [www.del.icio.us](http://www.del.icio.us).
6. Yahoo. *Flickr*. 2011 [retrieved: December, 2011]; Available from: [www.flickr.com](http://www.flickr.com).
7. Lamere, P., *Social Tagging and Music Information Retrieval*. Journal of New Music Research 2008. **37**(2): pp. 101-114.
8. Library of Congress, *Library of Congress Subject Headings*. 2011 [retrieved: December, 2011]; Available from: <http://www.loc.gov/aba/>.
9. European Patent Office, *European Patent Database*. 2011.
10. WIPO, *Strasbourg Agreement Concerning the International Patent Classification*, WIPO, Editor. 1971, World Intellectual Property Organization.
11. WIPO, *International Patent Classification (IPC)*, WIPO, Editor. 2011, World Intellectual Property Organization.
12. NISO, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. 2010, Bethesda, Maryland, U.S.A.: National Information Standards Organization Press.
13. Malone, T. *How do people organize their desks? Implications for the design of office information systems*. ACM Transactions on Information Systems (TOIS). ACM Press, 1983. **1**(1): pp. 99-112.
14. Lansdale, M., *The psychology of personal information management*. Applied Ergonomics, 1988. **19**(1): pp. 55-66.
15. Kidd, A. *The Marks are on the Knowledge Worker*. Proceedings of CHI on Human Factors in Computing Systems. 1994. Boston, USA: ACM: pp. 186-191.
16. Whittaker, S. and C. Sidner. *Email overload: exploring personal information management of email*. Proceedings of CHI'96. 1996. Vancouver, Canada: ACM: pp. 276-283.

17. Eysenck, M.W. and M.T. Keane, *Cognitive psychology: a student's handbook*. 5th Edition. 2005, Hove: Psychology.
18. Oren, E. *An Overview of Information Management and Knowledge Work Studies. Proceedings of the ISWC Workshop on the Semantic Desktop*. 2006.
19. Bowker, G.C. and S.L. Star, *Building Information Infrastructures for Social Worlds - The Role of Classifications and Standards*. Community Computing and Support Systems, 1998: pp. 231-248.
20. Hearst, M.A., *Clustering versus Faceted Categories for Information Exploration*. Communications of the ACM 2006. **49**(4): pp. 59-61.
21. Bondarenko, O. and R. Janssen. *Documents at hand: Learning from paper to improve digital technologies. Proceedings of CHI2005*. 2005. Portland, Oregon, USA: ACM: pp. 121-130.
22. Marsden, G. and D.E. Cairns. *Improving the Usability of the Hierarchical File System. Proceedings of the 2003 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on Enablement through Technology*. 2003. South African Institute for Computer Scientists and Information Technologists, Republic of South Africa: pp. 122-129.
23. Gifford, D., P. Jouvelot, M. Sheldon, and J. O'Toole. *Semantic File Systems. Proceedings of the 13th ACM Symposium on Operating Systems Principles*. 1991: ACM Press: **25**(5): pp. 16-25.

## Designing Multi-Modal Map-Based Interfaces for Disaster Management

Volker Paelke

3D Geo-Visualization and Modeling Group  
Institut de Geomàtica  
Barcelona, Spain  
volker.paelke@ideg.es

Christian Geiger

Department of Media Mixed Reality and Visualization  
Düsseldorf University of Applied Sciences  
Düsseldorf, Germany  
geiger@fh-duesseldorf.de

Karsten Nebe

Computer Sciences, Internet-Technologies and  
Usability Engineering  
Rhine-Waal University of Applied Sciences  
Kamp-Lintfort, Germany  
Karsten.Nebe@hochschule-rhein-waal.de

Florian Klomp maker, Holger Fischer

C-LAB  
University of Paderborn  
Paderborn, Germany  
{florian.klomp maker; holger.fischer}@c-lab.de

**Abstract**— The access to current and reliable maps and data is a critical factor in the management of disaster situations. Standard user interfaces are not well suited to provide this information to crisis managers. Especially in dynamic situations conventional cartographic displays and mouse based interaction techniques fail to address the need to review a situation rapidly and act on it as a team. The development of novel interaction techniques like multi-touch and tangible interaction in combination with large displays provides a promising base technology to provide crisis managers with an adequate overview of the situation and to share relevant information with other stakeholders in a collaborative setting. However, design expertise on the use of such techniques in interfaces for real-world applications is still very sparse. We are, therefore, conducting interdisciplinary research with a user and application centric focus to establish real-world requirements, to design new multi-modal mapping interfaces, and to validate them in disaster management applications. Initial results show that tangible and pen-based interaction are well suited to provide an intuitive and visible way to control who is changing data in a multi-user command and control interface.

**Keywords**— *post-WIMP user interfaces; natural user interfaces; mapping; geo-visualization; multi-touch interaction; pen-based interaction; tangible interaction*

### I. INTRODUCTION

Our goal is to improve the management of large-scale disaster situations and complex emergencies by providing crisis managers with an interactive mapping system that aims to improve their effectiveness. Crisis managers often lack access to vital information. With digital map repositories and the proliferation of new sensors the problem increasingly changes from one where information is missing to one where the required information is too difficult to access and

analyze. For example, even if the required information can be collected, crisis managers are still faced with the need to cope with large amounts of data that needs to be processed in order to guarantee successful operations.

Existing work on post-WIMP (Windows, Icons, Menu Pointer) user interfaces has largely focused on the development of base-technologies and individual visualization and interaction techniques. We complement this work using an approach that starts with the application and focuses on user centered design (UCD). In this process, we use close collaboration with end users to establish requirements, to examine new concepts in data analysis and presentation and to validate the newly developed user interfaces. With regards to user interface techniques, we are working both on the technology level and on the interface level. At the technology level we develop new multimodal interfaces, incorporating advanced techniques for interaction and information. At the interface level we provide interface solutions to real-world problems, specifically command and control interfaces with advanced multi-user capabilities.

In this paper, we initially review related work (Section II.) and then introduce our design approach (Section III.). The design of our multi-modal system for disaster management is detailed in the following sections, starting with the requirements (Section IV.). According to our hierarchical design approach these requirements are addressed by base technology components (Section V.) which are then combined into the complete system (section VI.). Our experiences and initial feedback from the end-users are reviewed in section VII. Finally, we draw conclusions from the design and evaluation process and provide an outlook on future work.

## II. RELATED WORK

Most relevant data in disaster management are of spatial nature (environment, location of resources) and one key challenge is to effectively integrate and unify spatial data from different sources. Geographic Information Systems (GIS) are well established to deal with the acquisition, storage, analysis and presentation of spatial data and are established tools in the management of crisis situations in many agencies. Challenges arise in the user interface, the integration of data and the integration with other systems. While the integration of static geo-data from different providers is addressed by initiatives like INSPIRE [4], a special challenge remains due to the fact that high-bandwidth sensors acquire much data in a disaster use-case at the time of use that is difficult to handle, analyze and present.

In addition, a crisis management system must also be able to handle different types of imprecise and non-digital data sources that arise in typical emergency situations.

A critical factor in the management of disaster situations is the access to current and reliable data. Novel sensors like infrared cameras, LIDAR [11] and SAR [3] allow to capture geo-spatial data when and where required, e.g., in the case of SAR irrespective of weather-conditions and visibility. Key challenges are the control of such sensors and the integration of such geo-spatial sensor information into a crisis management system. Especially in dynamic crisis situations, conventional displays interaction techniques fail to address the needs of crisis managers. Similarly, current GIS interfaces are not well adapted to use "in the field". The extension of GIS with novel techniques for real-time data handling and advanced interaction techniques is therefore required. In the recent years, multi-modal user interfaces and post-WIMP interfaces have seen a rapid development, especially the emergence of so-called natural user interfaces (NUI), sometimes referred to as reality-based interaction [12]. While many NUI techniques have been demonstrated in research, the design expertise on the use of such techniques in interfaces for large real-world applications like disaster management is still very sparse [7]. For the successful application of NUIs, it is therefore essential to integrate contributions from a number of different research areas within a user centered design process to adequately support users of disaster management systems.

A key aspect is an up-to-date and reliable presentation of the geo-spatial environment. In the past this has been mostly presented by (digital) maps, but advancing technologies in sensors, displays and interaction devices enable the integration of real-time data and the use of new displays and output devices to provide both mobile users (e.g., rescue forces) and decision makers with more adequate user interfaces. A central challenge is to adequately adapt these emerging technologies to geo-spatial information. Again it is possible to draw on previous research, e.g., in cartography, navigation and geo-visualization, but this has to be adapted, integrated and validated in the application context.

Some earlier research has been conducted in the field of disaster management that addresses technological support for collaboration and coordination. However, few systems

address the existing real-life-workflows of disaster management organizations like fire departments, police, medical services, etc. While interactive crisis management based on new interaction technology is regarded as promising [6], its adoption is limited because classical tools like pens and paper, paper-maps and plastic labels are proven and failsafe. A study of the potential use of multi-touch tabletops by the Dutch research institute TNO [13] investigated some possible uses of large scale interactive displays to provide effective assistance for decision-making during a disaster. It investigated in which departments such devices could be used, which workflows can be mapped and which not. The study shows that even casual users can use a tabletop without much learning effort and that it was effective in supporting collaborative work. The study also identified some problems in the specific system design. By using a user centered design approach from the beginning, we aim to avoid similar problems in our system.

## III. DESIGN APPROACH

A user centered approach is essential to develop new user interfaces that realize the benefits of technologies like post-WIMP user interfaces. In our project, we collaborate directly with the German Federal Agency for Technical Relief (Technisches Hilfswerk/THW) and draw on previous collaborations with fire fighters to adjust to real-world requirements [10]. To design and develop the system, we required an approach that takes both the requirements of large scale software engineering and usability engineering into account, while being adaptable to new and rapidly changing base-technologies on which little design expertise is available. Several approaches have aimed to integrate software engineering and usability engineering activities, either at the abstract overarching level of standards (serving as a framework to ensure consistency, compatibility, exchangeability, and quality), the level of process models (providing an organizational framework) or the operational process level (direct prescription of activities). For a detailed discussion see [9]. For our purposes the level of standards was too abstract to provide useful guidance and processes described at the operational level did not take the specific requirements (especially handling new and immature base-technologies) into account. Therefore, we started at the process model level, using the ISO standard 13407 (now replaced by 9241-210:2010) as our base, and derived an iterative operational process from this (Figure 1).

As our project involves stakeholders from many different disciplines, it was essential to provide an adequate mental model not only within the applications user interface (for the users), but also for the stakeholders in the development process (end-users, domain experts, experts for different base-technologies, designers, developers). We have, therefore, structured an iterative design process in a three level hierarchy (Fig. 1).

The base-technology level covers the hardware and software to enable an interaction or visualization, e.g., for a multi-touch device this would cover the sensors and associated processing to detect and process the touches of a user. In a project dealing with immature base-technologies it

is useful to decouple this development into a separate level, as many iterations or even changes in the base-technology employed (e.g., capacitive, resistive or optical multi-touch technologies) may be required.

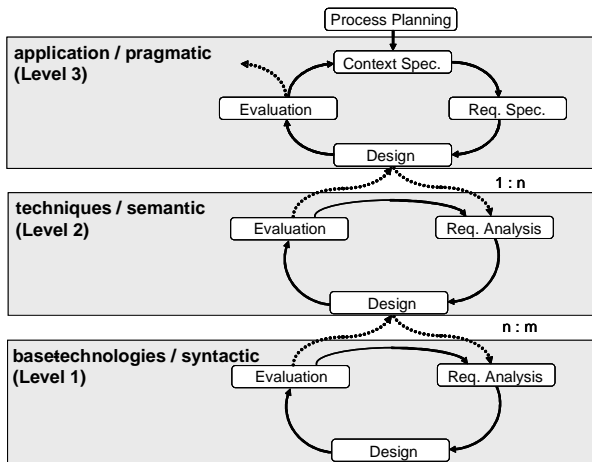


Figure 1. Hierarchical iterative design process based on ISO-13407

The second level covers individual interaction and visualization techniques. These employ the base-technologies to effect a useful task within the user interface, e.g., for a multi-touch device this could be the recognition of a specific touch-gesture. Again, it is useful to separate this design and development to achieve re-usable components. In mature environments this is typically not required because mature techniques are provided.

The third (application) level integrates different techniques from level 2, to construct a functional user interface for the application. A similar layering is applied to the data handling parts of the system.

Within the design process, the different activities correspond to common practice, specifically:

- Context of use
- User requirement (definition of scenarios, identification of requirements, definition of appropriate measures)
- Production of design solution (from scenarios over prototypes to implementation),
- Evaluation (analysis of requirements, review of designs, tests of prototypes, and evaluation of the complete system).

IV. REQUIREMENTS

The requirements elicitation and analysis is conducted as an iterative, on-going process. We were able to build on experience from previous projects with firefighters and the German Federal Agency for Technical Relief (THW). Existing studies on disaster management were also consulted to provide additional background information.



Figure 2. Observed THW exercises

However, the most important insights were derived from interviews, workshops and on-site training exercises conducted with disaster managers and technicians from THW. During training exercises we were able to observe the experts in their real work environment (Figure 2).

In order not to interfere with the exercises we installed video cameras and microphones in the operations center and recorded the course of events during the exercise. Based on the recordings we were able to analyze the details of each workflow. These insights were captured in scenarios that cover the different types of crisis situations and their information and command requirements, user roles (including professionals, volunteers, local collaborators, politicians, press, public), and operating environments (fixed and mobile command centers, operations in disaster zones).

A typical scenario is the management of a flooding incident by a THW team. Tools used include paper-based maps and special sheets, so called ‘damage accounts’, that contain summaries of local incidents and assigned units. Sheets are referenced to map locations by using little magnetic tiles. Typical activities are the creation, manipulation and spatial update of damage accounts.

These scenarios were then analyzed, to derive key requirements at the application, task and interaction level.

General application level requirements include:

- Provide access and control to information the way users are used to
- Manage and visualize the current situation in the field
- Maintain the benefits of the established robust workflow, that is clearly visible to all stakeholders
- Easily integrate non-expert personal (e.g. local support staff)
- Clear allocation of control for critical tasks
- Support for information sharing
- Separation of situation display and planning
- Data interface with OGC standard and commonly used non-standard data formats
- Enable the integration of imprecise and non-digital data sources
- Provide understandable presentations for different user groups (expert, local collaborator)
- Enable fast and easy communication and sending of commands to mobile units in the field
- Enable integration of software tools that allow a more efficient processing of recurring tasks

Expected additional benefits for a new system include:

- Seamless and scalable map display
- Support for rich media presentations of information
- Selective use of information layers
- Support for geo-referencing of units and incidents and automated transmission of coordinates
- Integration with existing GIS systems
- Access to real-time sensor and location data
- Information filtering and spatial analysis functions
- Support for private workspaces
- Combination of co-located and off-site interaction
- Ability to distinguish between different users; traceable interaction

The requirements analysis revealed a number of areas for potential improvements and indicated the need for a collaborative visualization and interaction environment. Natural user interface technologies like multi-touch, tangible, gestural and vocal interaction seems promising, but their suitability and usability must be confirmed. To ensure that the interaction remains coherent and “natural” the design of the system must ensure that users develop and maintain a suitable mental model of the system in operation.

#### V. BASE-TECHNOLOGIES AND TECHNIQUES



Figure 3. The useTable and its interaction possibilities (pen, multi-touch, tangible puck)

As described previously, we address the design and development at three hierarchical levels. The separate consideration of the base-technology and interaction / visualization technique level allows to change base-technologies during development (if required) and to develop interaction and visualization techniques that can be reused.

As the central hardware component for interaction and visualization we build on the ‘useTable’ [14], a flexible visualization table constructed at C-LAB, that supports multi-touch, tangible and pen-based interaction. Compared to off-the-shelf solutions this approach enables us to adapt the technologies and techniques to the application and does not limit the design to the constraints of a given hardware environment. Using the feedback collected throughout the design process, the useTable has evolved into an interaction environment adapted to disaster management requirements.

The ‘useTable’ consists of a 55” display that offers full HD image projection. The projector is mounted beneath the

surface and the image is projected on the top surface by two mirrors. For finger-tracking FTIR (Frustrated Total Internal Reflection) is applied and objects on the surface are tracked using combined DI (Diffused Illumination). The camera on the bottom of the table is equipped with a corresponding IR filter and is connected to a tracking PC that applies the filter and tracking algorithms. The projection surface is equipped with an antireflex diffusor sheet that enables pen-based interaction by using Anoto digital pens [1, 2].

On the software side, a new detection and tracking framework for advanced interaction using a depth-sensing camera [6], called dSensingNI, was developed. The dSensingNI framework is capable of tracking user fingers and palm of hands, which enables precise and advanced multi-touch interactions as well as complex tangible interactions. For tangible interaction arbitrary physical objects can be used to control interaction. Using the depth-sensing camera, physical objects can be used in common (2D) actions, such as placing and moving, and also in 3D actions, such as grouping or stacking. The depth-sensing also allows extending the multi-touch interaction to object surfaces without the need for integrated logic and sensors.

Combing RFID chips and depth-sensing cameras we are able to identify and track the persons that are interacting with the useTable. This allows applying different functionality to different users based on their roles during the interaction, a central requirement not addressed by off-the-shelf multi-touch tables.

Using the useTable and dSensingNI as base technologies, a number of different interaction and visualization techniques have been implemented. These techniques enable experiments with users, e.g., to study the usability differences between touch input, pen-input and the use of interaction-objects. A key advantage of the interactive display in the disaster management application is the ability to rapidly switch between different maps and map representations. Using a layer concept different maps and additional information (e.g., airborne imagery) can be mixed while maintaining the established workflow. The extension of the visualization beyond map-display allows experimenting with integration of derived information (e.g., danger zones, uncertainty) as well as task dependent map generalization and highlighting strategies.

Insights from these studies are used to guide the development at the base technology level. For example, experience showed that in some scenarios a strict separation between visualization of the current situation and the planning of future actions is essential. Our design approach allows to adapt to these requirements by modifying and extending the set of available base technologies. To provide an intuitive separation we extended the useTable into an L-Shape display. The L-Shape employs the useTable for planning as described. An additional wall display was added to visualize the current situation.

## VI. SYSTEM

Building on the base technologies and interaction techniques we have implemented various iterations of a functional prototype that covers the functionality required by the THW to handle a regional flooding incident.

The system integrates, analyzes and presents the spatial data pertaining to the situation. To achieve this, it integrates digital maps, air- and space-borne imagery and a 3D terrain model. The use of a 3D terrain model enables important interactive analysis functions not available in the traditional setup using paper-maps, e.g., the calculation of number of pumps required to transport water along a specific trajectory.

According to the layering concept the system is structured into three different levels.

Essential tasks at the application level to be supported in this use-case are the communication and update of the current situation in the field (requirements: display of maps and additional information layers; editing of damage accounts), the planning of future actions and assignation of units (information input; communication). These in turn require appropriate interaction and visualization techniques.

For the visualization we started with a digital equivalent of the forms, signs and labels that are standardized and familiar to the THW staff. Additional features not present in the conventional environment include the possibility to overlay additional (geo-referenced) information layers, dynamic changes of symbolization and the level of detail presented, as well as the possibility to zoom, pan and rotate the map (Figure 3).

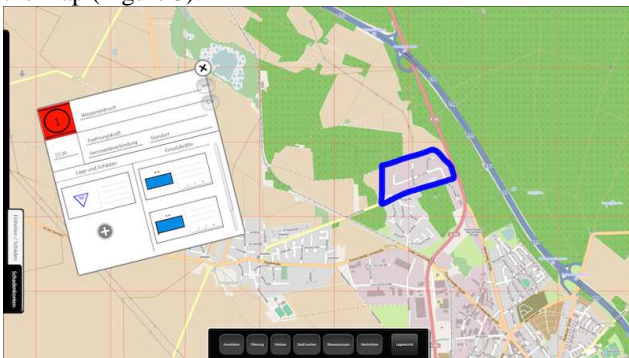


Figure 4. Situation overview with damage account

On the interaction side techniques were required to control the system and visualization (system control), as well to conduct primary tasks, e.g., the creation and update of damage account documents that capture the current situation, and the assignment of vehicles and entities to different damages in the field.

Early feedback indicated, that for editing damage accounts the use of a digital pen (mimicking the traditional paper forms) was regarded as preferable to traditional hardware keyboard or on-screen software keyboard input (Figure 4).



Figure 5. Editing damage accounts

In practice sessions, the digital pen also proved to be more suitable for marking and planning tasks than finger-touch input, as it allows more precise input.

While multi-touch gestures for scrolling, zooming and rotating have become popular with smartphones and tablets, early tests revealed that these are not suitable in our application scenario: Multiple users touching simultaneously (e.g., to comment) can easily lead to undesired changes. In the disaster management application usually one person should be in charge of the map representation, e.g., when changing the data layers displayed or the map scale.

Tangible interaction ([12]) using a physical puck (see Figure 2, bottom right) provides a suitable interaction technique: The puck is placed on the useTable surface - by moving and rotating the puck the map can be translated and zoomed. Since there is only one puck, it is always clear to all collaborators who is currently controlling the map. This considerably enforces group and interaction awareness.

Since all information is available in digital form the system enables simulation and planning capabilities, that are not available in the conventional workflow. E.g., in planning the transport of water between two locations the systems provides support to calculate the number of pipe sections and the number of pumps required (using a digital terrain model for the calculation). The system improves on the state of the art at all three levels - at the syntactic level by extending the scope of geo-spatial data-sources from static maps towards a wide range of (dynamic) data sources, at the semantic level by providing analysis function and at the pragmatic level through a geo-visualization component that exploits the benefits of post-WIMP interaction techniques.

## VII. EXPERIENCES AND FEEDBACK

As explained in section 3 the ongoing development takes place in close collaboration with the intended end-users from THW. In addition to formative evaluation that guides the development we have also conducted initial tests with experts from the THW and also discussed the system with members of the THW authority. The feedback has been very positive. Even small improvements enabled by the digital map (e.g., switching maps while keeping the data and

annotations geo-referenced) caused enthusiastic responses. Process improvements enabled by having all data in digital form (e.g., the calculation of the number of required pumps) lead to significant improvements in efficiency (in the example of calculating water pipelines a simply drawing of the intended connection with immediate feedback replaces a manual process that required 30min with paper maps and required significant experience to avoid calculation errors).

Initial feedback also led to a number of interesting insights into post-WIMP interaction techniques. E.g., while multi-touch gestures for rotation and translation are well established and one of the typical showcases for multi-touch interaction, we found out that they are not applicable in a mission critical multi-user map application. This is due to the potential for un-intended and non-comprehensible transformations and the need for traceable commands. For other interactions (especially in temporary local workspaces) multi-touch gestures were found suitable. The experience with digital pens and tangible indicates that they enable very natural interactions in our application context with correspondingly high acceptance by users.

In future work, we aim to complement the formative evaluations with more comprehensive user studies and tests, to study the usability of different interaction and visualization techniques. In addition, we also aim to study the impact on the cognitive workload of users and examine the potential physiological dangers that may be incurred by prolonged use of a large-scale post-WIMP display. While the ergonomic requirements of desktop workplaces are well understood, the same is not true for new interaction environments like the useTable. Established ergonomic standards were often ignored in early demonstrators of post-WIMP techniques because of technology constraints (e.g., by limiting lighting levels to reduce IR contamination). And while interaction techniques like free-hand gestures are intuitive they can also cause a high-level of fatigue.

### VIII. CONCLUSION

Research is required to exploit the potential of advanced visualization techniques and user interaction techniques in disaster management and similar applications. The step from technology demonstrators to usable real-world systems requires adequate tools as well as stable base-technologies and the evaluation and validation of different design options. In this paper we present an initial step in this direction with a focus on user centered design and development in a disaster management application.

A promising extension of the system would be to extend the support not only to the planning staff in the command center, but also to individual rescue workers in the field. In a separate project (FireNet, [16]), we have conducted early experiments with a mobile personal sensor network, in which each rescue worker was equipped with a sensor node (using Sun's SunSpot as the basis and extending it with GPS receiver and additional sensors). Integrating such functionality within the disaster management application could further improve situation awareness in the command center by allowing real-time tracking of rescue personal as well as equipment. Another area for future work concerns

the extension from the current focus on observation (situation awareness) to analysis and prediction. The digital data enable the use of analysis functions (as exemplified by the pump planning) and a future extension towards simulation/prediction could be useful, especially in dynamic natural disaster situations like flooding or fires.

### ACKNOWLEDGMENTS

The authors wish to thank the German Federal Agency for Technical Relief (Technisches Hilfswerk/THW) in Detmold for the productive collaboration, specifically Torsten Meier and Oliver Charles as well as the students in our students project groups at the University of Paderborn. We would also like to thank Petre Dini for his helpful comments during the revision of the paper.

### REFERENCES

- [1] ANOTO: <http://www.anoto.com>; (last accessed 11.01.2012).
- [2] Haller, M., Brandl, P., Leitner, J., and Seifried, T.: Large interactive surfaces based on digital pens. In: 10th International Conference on Humans and Computers, pp. 172-177 (2007).
- [3] Hanssen, R.: Radar Interferometry: Data Interpretation and Error Analysis, Kluwer Academic, 2001.
- [4] Inspire Directive (Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007): <http://inspire.jrc.ec.europa.eu/> (last accessed 11.01.2012).
- [5] Jokela, T.: An Assessment Approach for User-Centred Design Processes. In: Proceedings of EuroSPI 2001. Limerick: Limerick Institute of Technology Press (2001).
- [6] Jung, H, Nebe, K., Klompaker, F., and Fischer, H.: Authentifizierte Eingaben auf Multitouchischen, Proc. Mensch und Computer, Chemnitz, Germany, 2011, pp. 305-309.
- [7] Kobayashi, K., Kakizaki, T., Narita, A., Hirano, M., and Kase, I.: Tangible user interface for supporting disaster education. In: Proceedings of SIGGRAPH '07, ACM, New York (2007).
- [8] Leitner, J., Powell, J., Brandl, P., Seifried, Th., Haller, M., Dorray, B., and To, P.: Flux: a tilting multi-touch and pen based surface. In: CHI '09 Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, ACM, New York (2009), pp. 3211--3216.
- [9] Nebe, K.: Integration von Usability Engineering und Software Engineering, Dissertation, University of Paderborn, Shaker, 2009.
- [10] Nebe, K., Müller, T., and Klompaker, F.: An Investigation on Requirements for Co-located Group-Work using Multitouch-, Pen-based- and Tangible-Interaction. In: Proceedings of the HCI2011, Springer-Verlag, New York, 2011, pp. 90-99.
- [11] Shan, J. and Toth, C. (Eds.): LIDAR: Topographic Laser Ranging and Scanning: Principles and Processing, CRC Press, 2008.
- [12] Terrenghi, L., Kirk, D., Richter, H., Krämer, S., Hilliges, O., and Butz, A.: Physical handles at the interactive surface: Exploring tangibility and its benefits. In: Proceedings of the working conference on Advanced visual interfaces, ACM, New York (2008), pp. 138—145.
- [13] TNO: <http://www.tno.nl/> (last accessed 11.1.2012).
- [14] UseTable: <http://www.usetable.de/>; (last accessed 11.1.2012).
- [15] van Oosterom, P.; Zlatanova, S.; and Fendel, E. (Eds.): Geo-Information for Disaster Management, Springer, 2005.
- [16] <http://www.ikg.uni-hannover.de/index.php?id=602&type=98>



# Multilingual Ontology Alignment Based on Visual Representations of Ontology Concepts

Srdan Mihic and Dragan Ivetic  
 Computing and Control Department,  
 Faculty of Technical Sciences, University of Novi Sad  
 Novi Sad, Serbia  
 e-mail: {smihic, ivetic}@uns.ac.rs

**Abstract**— Image search represents one of the most frequent user actions on the Internet. Existing image search engines do not understand the images they return, nor do they support multilingualism. These issues can be addressed with the introduction of a semantic layer. The semantics is encoded in ontologies, which contain structured information about a domain of application. In order to provide semantic interoperability between (multilingual) ontologies, it is necessary to obtain semantic correspondences – ontology alignments. Various strategies have been proposed for multilingual ontology alignment. In this concept paper, the idea of alignment discovery based on semantic similarity of visual representations of ontology concept is explored.

**Keywords**-multilingual ontologies; ontology alignment; image retrieval; multimedia semantics

## I. INTRODUCTION

People often use Internet for querying images. Existing image search engines are syntax-based, and thus do not understand the images they return. The result sets are mostly large but lack precision. Namely, a good part of the result set is irrelevant to the formulated query. Introduction of a semantic layer in image retrieval improves the precision of results obtained [1].

Current image search engines have very limited support for multilingualism. Although they provide users with ability to narrow the region and/or language (used for image tagging/description) but that does not provide satisfactory results. In the following, some of the pressing issues will be presented.

Distribution of images in relation to languages is non-uniform on the Internet. Usually, the higher the presence of a language, the bigger result set is retrieved. For example, a query for 'шак' (Serbian for 'sack') produces no semantically valid results. Yet issuing equivalent query in English produces a vast number of semantically valid results with high precision.

In linguistics, homographs are group of words that share the same spelling but have different meanings, regardless of how they are pronounced. Homonyms are homographs that have the same pronunciation as well. Word in one language is often a homograph/homonym for an unrelated word in another language – a cross-lingual homograph/homonym. Therefore, it is possible for one word (in language with higher presence) to mask the other (in language with lower presence). For example, a query for 'fog' (Hungarian for 'tooth') yields in images of misty weather (because higher presence of English than Hungarian).

User queries are often composed of a few words (generally two or three words), and are too imprecise to express the query that the user had in mind [2]. Especially, it is hard to formulate proper queries in image search [2]. In addition, studies have shown that users tend to look only at the first answers pages [3]. Thus, it is necessary to obtain and rank the “right” answers first based on a short fuzzy description.

Images, that are relevant to the formulated query, are retrieved if a user queries in the “right” language. Thus, users have to issue queries in various natural languages in order to obtain satisfactory results. Not all users have necessary linguistic skills to adequately translate queries in a foreign language. Even translation tools fail to provide adequate translations. This results in imprecise translations that can lead to even poorer set of results. For example, the aforementioned term 'шак' can be translated as 'bag'. This translation is more imprecise than one with the term 'sack', but more common for users who do not know English language well. As expected, this translation yields no satisfactory results. Therefore, automatic inclusion of translation of terms in a semantically meaningful way would provide a richer set of retrieved images and would lead to an enhanced user experience.

By addressing the aforementioned issues users would be able to state queries in the language of their choice and to get the most appropriate image results regardless of the language used.

Ontologies represent an economic and efficient way to address aforementioned issues and to model semantic layer. Thus, in recent years, they have gained a large amount of attention and many have been developed and are available on-line.

With the expansion of ontologies in terms of application domains, the number of natural languages in which they were written grew. Thus, reasoning and mapping of these multilingual ontologies has become an important issue [4].

The process of linking related ontology elements is called ontology alignment (or mapping) [5][6]. Ontology alignment enables semantic interoperability between distributed information systems. The resulting alignments can be used for agent communication (interoperability between distributed information systems), query answering (executing query in all available natural languages), ontology merging, or for navigation on the Semantic Web [7].

There are many ontology alignment techniques (see [6] for an exhaustive review) and various multilingual ontology alignment strategies have been proposed (see Section V for detailed review). Common to all these solutions is that the

generation of alignments is based on comparison of (multilingual) ontology labels.

In addition to these approaches, we propose to use images as visual representations of ontology concepts for alignment discovery between two multilingual ontologies. This approach complements the aforementioned approaches.

The outline of this paper is as follows. In the following section, the image-based multilingual ontology alignment approach for building indirect mapping between multilingual ontologies is described as the main contribution of this work. In Section 3, the initial proposal for image-based alignment discovery is presented. In Section 4, some previous studies related to this work are introduced. Finally, Section 5 concludes this paper and presents further research directions.

## II. IMAGE-BASED MULTILINGUAL ONTOLOGY ALIGNMENT

We draw our inspiration from the natural way the humans learn new languages. One can learn a foreign language visually by establishing pictorial inter-language mappings between visual representations of corresponding terms/concepts. These pictorial inter-language mappings have proven quite useful in a number of commercial language learning applications, as for instance Rosetta Stone [16], and therefore applying it more formally to ontology alignment might be a promising idea.

For example, let us consider a situation in which two speakers want to communicate with each other (Fig.1). The first speaker is from Serbia and speaks only Serbian, and the other one is from Japan and speaks only Japanese. Unfortunately, neither of them knows the language spoken by the other one, nor they speak the common language. If they want to communicate with each other, they will have to teach each other their respective languages. The most natural way to this is to use real life objects, more precisely their visual representations (images), and to exchange their labels (in Fig.1 using image of a dog the speakers learn its label in foreign language). This way, the speakers will most likely learn the most common and the most adequate word meaning.

In many natural languages, entities are described by nouns, which are, in majority, picturable entities [1]. The number of nouns in natural language is usually significantly higher than

the number of verbs, adjectives, and adverbs (e.g., 80% of the Serbian WordNet are nouns [9]). As building blocks of ontologies, concepts and their instances are described by nouns as well. Espinoza et al. [10] have empirically found that existing ontologies share the same lexical patterns. For example, approximately 60% of concept labels follow an adjective-noun pattern (e.g., *temporal region*), where as the others (about 30%) use the noun-noun pattern (e.g., *knowledge domain*) [10]. We limit our discussion to the above stated lexical patterns. Other lexical categories (e.g., verbs) are left for future research, since some of them can be represented by picturable entities as well.

Cognitive psychology studies have found that: i) there exists a correlation between visual and semantic similarity in the human visual system; ii) semantic categories are visually separable; iii) there exist visual prototypes for semantic categories [11]. More recently, Deselaers et al. [12] have experimentally confirmed that these conclusions hold in the field of computer vision. In addition, they have found that the visual variability within a category grows with its semantic domain.

There are plenty of images available on-line that can be used as visual representations of ontology concepts.

According to the aforementioned, visual representations of ontology concepts can be used and compared in order to find out the adequate mapping. Our idea is additionally supported by the fact that it is easy to cope with synonyms issues in visual domain since synonyms visual representations are similar or even the same (e.g., words *hound* and *dog* are synonyms and visually represent the same entity).

A proposed architecture for image-based multilingual ontology alignment is presented in the following section.

## III. A SKETCH OF A POSSIBLE SYSTEM ARCHITECTURE

The proposed architecture is shown in Fig. 2. It is based on four main components: the Alignment Generator, the Visual Representations Provider, the Image Comparator, and the Alignment Repository.

The Alignment Generator receives two ontologies as input and generates alignments if possible. Firstly, for each concept pair of the matching ontologies, the component checks whether suitable alignment already exists in the Alignments Repository. If it does not, this component enquires the Visual Representations Provider to provide suitable visual representations (several images and their accompanied textual descriptions) of these concepts. If such representations can be found, they are compared using the Image Comparator component. The Image Comparator computes a degree in which these visual representations of concepts are related and chooses the best among these representations. Finally, the alignment is generated and stored in the Alignments Repository for sharing and reuse, along with the chosen visual representations.

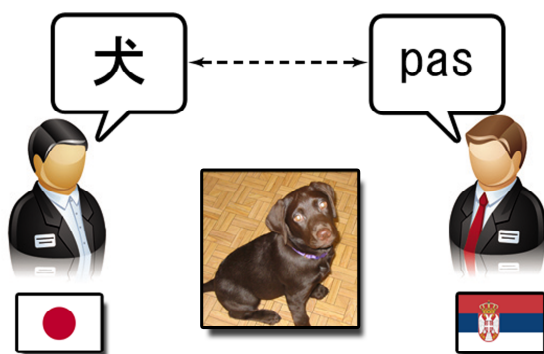


Figure 1. Natural way of learning terms of foreign language using its visual representation

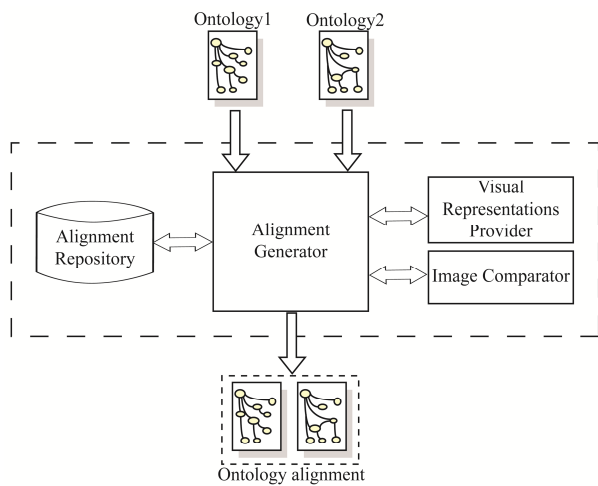


Figure 2. The proposed architecture for image-based multilingual ontology alignment

Later, the generated alignments can be employed for automatic query translation and distributed query answering.

The proposed system is in the early stages of development and provides guidelines for future work.

A. Acquisition of the visual representations of concepts

The process of multilingual ontology alignment discovery begins with a comparison of leaf concepts. The rationale for comparing leaf concepts first is that they often point towards specific entities, while concepts that are high in the hierarchy tend to represent more abstract and thus more ambiguous entities [1]. In addition, for those concepts the semantic domain is narrow, thus visual variability is small (see Section II).

For each concept pair, the Visual Representations Provider component tries to provide suitable visual representations of these concepts, as well the accompanying text (tags, annotations, etc.). The success of the entire process of image-based alignment discovery is highly dependent upon this step. The image comparison process is more reliable and precise if the acquired images are true semantic visual representatives of concepts.

Thus, the component should attempt to acquire images from semantically rich sources, if possible. The component attempts to find the source that supports queries stated in the natural languages of both ontologies first. If no such source can be found or yields no results, the component opts for two monolingual sources: one for each natural language. If those sources cannot be found or yield no results, the component opts for sources with less support for semantic search. Four types of sources have been identified according to the semantics they incorporate: ontology-based image retrieval systems, hierarchy-based image databases (usually WordNet [13] hierarchy-based), content-based image retrieval systems, and syntax-based image search engines. The sources are listed in descending order of the semantics they incorporate.

In situations where none of aforementioned steps obtain visual representations, the system marks that pair as incomparable due to lack of data and steps to the next pair.

When querying for images, the context of the ontology concept is used to disambiguate the lexical meaning of a

concept label, as in [10]. For example, let us consider an ambiguous concept label *crane*. The term *crane* can have two senses in English: *a bird* and *a type of construction equipment*. Thus, an image search with the term *crane* results in images both of birds and of construction equipment. By adding parent concept label/s (e.g., *bird*) to the query, the obtained images are more appropriate than those using the concept label alone.

B. Semantic-based image comparison

After obtaining visual representations, the Image Comparator component performs semantic-based comparison of two visual representation sets and selects the best visual representations. This component is the core of the system and represents the most complex part of the system. For the time being, this component is in early stages of development. We plan to develop it as a multimodal probabilistic framework, inspired by [14, 15].

C. Generation of alignments

When computing, the confidence value reliability of the source must be taken into account. Source reliability is a weighting factor ranging from 0 to 1 which is used to define the influence of a particular retrieval option on the final result. Generally speaking, ontology-based retrieval is assigned high values, and syntax-based low values due to the greater reliability of the former.

If the confidence value is below a predefined threshold *t*, the concepts are considered unrelated. Otherwise, an alignment with a calculated confidence value is generated. In addition, to support alignment reuse, the algorithm stores alignments and respective visual representations in a shared alignment repository, similar to [16].

IV. USAGE SCENARIO

One possible usage scenario would be to use the generated alignments in the Alignments Repository to support automatic query translation into several natural languages.

For example, a Serbian teacher gives an assignment to her/his pupils, still in elementary school, to write an essay about the culture of modern Japan for a sociology class. Since pupils are not fluent in English nor do they know Japanese, it is very hard for them to acquire materials (including images) using common image search engines. First, they must face the problem of query translation in English and/or in Japanese. They do either this manually or by using some (machine) translation tools. It is highly unlikely that this approach would lead to acceptable result set. Secondary, they have to manually issue queries in both languages and compare them manually.

When relying on our approach, pupils can issue queries in their native language without the need to know any ‘major’ language. The query is parsed and concepts and context are extracted. The Alignment Repository is queried for those concepts. If alignments can be found, the concepts are translated in their respective equivalents in different languages. Since the alignments store the image data, which are visual representations of those concepts, these images can be used either as results and/or to support query-by-semantic-example [17] queries. The image search engines execute translated queries. The results are aggregated and presented to the user.

## V. RELATED WORK

Various multilingual ontology alignment strategies have been proposed: manual processing, the corpus-based approach, the linguistic enrichment, and the two-step generic approach [4].

Laing et al. [18] used manual mapping to map agricultural thesaurus in English to the Chinese equivalent. Whilst the manual mapping is costly and error-prone task, this approach is feasible only for relatively small and simple ontologies. Thus, fully/semi-automated multilingual ontology mapping strategies have emerged.

Corpus-based approaches use bilingual corpora for alignment discovery. In [19], by using this approach, Dutch thesaurus is aligned with the English thesaurus WordNet. This approach is applicable in situations where corpora of similar granularity and quality exist. Alas, for many domain-specific ontologies there are no adequate corpora to be used. In addition, this approach does not consider structural aspect and thus cannot provide precise mappings for ontologies with complex structure [4].

In instance-based approach, analysis of instance similarity is used for obtaining matching correspondences. This approach is based on machine learning methods and thus, is applicable for ontologies with sufficiently large number of instances. In [20], Wang et al. used annotations of instances to compute a measure of similarity between instances. Later this similarity was used to determine similarity between concepts.

According to proponents of linguistically enrichment strategy current ontologies suffer from unreadability due to badly chosen labels, lexical ambiguity etc., and thus, impeding the interoperability. They enrich the ontology's linguistic expressivity, through the exploitation of existing linguistic resources. A linguistically motivated mapping method has been proposed in [21]. Although linguistically enrichment of ontologies is beneficial, it is difficult to apply due to lack of linguistic resource standards.

In the generic two-step method, which was proposed in [7], the source ontology labels are translated into target language first and then monolingual matching techniques are applied. Since the translation does not take into account the semantics of involved ontologies, it can introduce inadequate translations that hamper the matching process. In these systems, the translation phase is crucial to success of ontology alignment. Therefore, obtaining the most suitable label translation is the key to generation of high quality alignments [4]. Fu et al. [4] addressed these issues with appropriate translation selection component. Namely, this component chooses the most appropriate translation amongst candidates with regard to target ontology semantics, the mapping intent, the operating domain, the time and resource constraints and user feedback.

Still, none of these approaches presents a comprehensive solution to the multilingual ontology alignment problem [22]. Thus, the multistrategy approaches have emerged. Various papers report that combination of strategies is highly dependent of the ontologies used. In [22], Li et al. presented a dynamic multistrategy ontology framework. They have used various similarity factors to select dynamically the most appropriate strategy for each individual alignment task. On the other hand, Songyun et al. propose an iterative supervised-learning weighted multistrategy alignment approach [23]. For each

alignment task, system computes weights for every strategy available and uses those weights to combine correctly the strategies.

We propose a conceptual idea to use images for alignment discovery between two multilingual ontologies. Unlike previous approaches images are used (visual representations of ontology concepts) to perform alignment discovery. Our idea is based on the fact that majority of ontology concepts are picturable entities, which can be found on the Web as images. Our approach complements the aforementioned approaches and adds a new dimension to the research field of multilingual ontology alignment.

## VI. CONCLUSION AND FUTURE WORK

This paper is a concept paper, which introduced the idea of indirect alignment between multilingual ontologies by discovering alignments based on semantic-similarity of visual representations of ontology's concepts. Thus, the problem of finding adequate alignment between two concepts is reduced to the problem of matching their visual representations.

To the best of our knowledge, this is the first time the images as visual representations of ontology concepts, are exploited for multilingual ontology alignment.

Our idea is appealing, but has following limitations:

- The approach is suitable in situations where visual representations of concept exist and are available. The more the concept is visually discriminating, the easier is to obtain alignment using image similarity, and vice versa. For some broad and abstract concepts, the approach is not feasible because of their visual diversity (e.g., *animal* concept has very broad visual diversity). For some others concepts no appropriate image/s can be found on the Web.
- Comparing two images is complex, computationally expensive and context-dependant task itself.

As future work, we want to: i) conduct experiments for evaluation of proposed idea and level of applicability; ii) implement a prototype that is build upon the presented idea; iii) investigate how this idea could be combined with existing strategies into synergy-based dynamic multistrategy alignment framework to enhance alignments accuracy and precision.

## ACKNOWLEDGMENT

This research is financial supported by Ministry of Science and Technological Development, Republic of Serbia; under the project number III47003 "Infrastructure for Technology Enhanced Learning in Serbia", 2011-2014.

## REFERENCES

- [1] A. Popescu, "Image retrieval using a multilingual ontology". In Large Scale Semantic Access to Content (Text, Image, Video, and Sound) (RIAO '07). Le Centre De Hautes Etudes Internationales D'informatique Documentaire, Paris, France, pp. 461-474 (2007)
- [2] S. Wang, F. Jing, J. He, Q. Du, and L. Zhang, "IGroup: presenting web image search results in semantic clusters". In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07). ACM, New York, NY, USA, pp. 587-596. doi:10.1145/1240624.1240718 (2007)
- [3] B.J. Jansen, A. Spink, and J. Pedersen, "An Analysis of Multimedia Searching on Altavista". In Proc. of the 5<sup>th</sup> ACM SIGMM MIR Workshop, Berkeley, CA, USA, (2004).

- [4] B. Fu, R. Brennan, and D. O'Sullivan, "Cross-Lingual Ontology Mapping and Its Use on the Multilingual Semantic Web". In 1st Workshop on the Multilingual Semantic Web, at the 19th International World Wide Web Conference (WWW 2010), Raleigh, USA, April 27th, 2010. CEUR Vol.571, 13 – 20 (2010)
- [5] E. Beisswanger, "Exploiting relation extraction for ontology alignment". In: Patel-Schneider, P. F., Pan, Y., Hitzler, P., Mika, P., Zhang, L. (eds.) ISWC'10, LCNS, Vol. 6497/2010 Part II, pp. 289-296, Springer, Heidelberg (2010)
- [6] J. Euzenat, and P. Shvaiko, *Ontology matching*. Springer, Heidelberg, Germany (2007)
- [7] C. Trojahn, P. Quaresma, and R. Vieira, "A framework for multilingual ontology mapping". In: Proceedings of the International Conference on Language Resources and Evaluation, LREC. pp. 1034- 1037 (2008)
- [8] Rosetta Stone, a language-learning software, [Online] Available at: <http://www.rosstattstone.co.uk>, 11/28/2011
- [9] S. Antonic, and C. Krstev, "Serbian Wordnet for Biomedical Sciences". In INFORUM 2008: 14th Conference on Professional Information Resources, Prague, May 28-30. (2008)
- [10] M. Espinoza, A. Gomez-Perez, and E. Mena. "Enriching an ontology with multilingual information". In Proceedings of the 5th European semantic web conference on The semantic web: research and applications (ESWC'08), S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis (Eds.). Springer-Verlag, Berlin, Heidelberg, 333-347. (2008)
- [11] R. J. Sternberg. *Cognitive Psychology*. Wadsworth, 5th edition, 2008.
- [12] T. Deselaers and V. Ferrari, "Visual and semantic similarity in ImageNet", in CVPR, IEEE, , pp. 1777-1784 (2011)
- [13] WordNet, a lexical database for the English language, [Online] Available at: <http://wordnet.princeton.edu>, 11/28/2011
- [14] R. Paredes and F.M. Segarra, "Interactive Image Retrieval". *Multimodal Interactive Pattern Recognition and Applications*. A. H. Toselli, E. Vidal and F. Casacuberta (Eds.), Springer (2011)
- [15] J. Milgram, R. Sabourin and M. Chretien, "Combining Model-based and Discriminative Approaches in a Modular Two-Stage Classification System: Application to Isolated Handwritten Digit Recognition", *Electronic Letters on Computer Vision and Image Analysis*, Volume 5, Issue 1, pp. 14-29 (2005)
- [16] Jung, J.J., Hakansson, A., Hartung, R.: Indirect Alignment between Multilingual Ontologies: A Case Study of Korean and Swedish Ontologies. In: Hakansson, A., Nguyen, N.T., Hartung, R.L, Howlett, R.J, Jain, L.C. (eds.) KES-AMSTA '09, LCNS, Vol. 5559/2009, pp. 233-241, Springer, Heidelberg (2009)
- [17] E. G. K. Chang, G. Sychay and W. Gang, "CBSA: content-based soft annotation for multimodal image retrieval using bayes point machines". *IEEE Transactions on Circuits and Systems for Video Technology* 13(1) pp. 26–38 (2003)
- [18] A. C. Liang, and M. Sini, "Mapping AGROVOC and the Chinese agricultural thesaurus: definitions, tools, procedures". *New Review of Hypermedia and Multimedia*, vol. 12:1, pp. 51-62 (2006)
- [19] V. Malaise, A. Isaac, L. Gazendam, H. Brugman, "Anchoring Dutch Cultural Heritage Thesauri to WordNet: Two Case Studies". In: Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007), pp. 57 – 64, Prague, Czech Republic (2007)
- [20] S. Wang, G. Englebienne, S. Schlobach, "Learning concept mappings from instance similarity". In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T.W. Thirunarayan, K. (eds.) *The Semantic Web - ISWC 2008, 7th International Semantic Web Conference, ISWC 2008, Karlsruhe, Germany, October 26-30, 2008, LCNS, vol. 5318, pp. 339-355 (2008)*
- [21] T. M. Paziienza, A. Stellato, "Linguistically Motivated Ontology Mapping for the Semantic Web". In: Proceedings of SWAP 2005, the 2nd Italian Semantic Web Workshop, Trento, Italy, December 14-16, 2005, CEUR Workshop Proceedings, vol. 166, (2005)
- [22] J. Li, J. Tang, Y. Li, Q. Luo, "RiMOM: A Dynamic Multistrategy Ontology Alignment Framework Journal", *IEEE Transactions on Knowledge and Data Engineering - TKDE* , vol. 21, no. 8, pp. 1218-1232 (2009)
- [23] D. Songyun, F. Achille, S. Kavitha, "One Size Does Not Fit All: Customizing Ontology Alignment Using User Feedback", In: Patel-Schneider, P. F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J., Horrocks, I., Glimm, B. (eds.) ISWC'10, LCNS, vol. 6496. pp. 177-192. (2010)

# TsoKaDo: An Image Search Engine Performing Recursive Query Recommendation Based on Visual Information

Lazaros T. Tsochatzidis Athanasios Ch. Kapoutsis Nikos I. Dourvas  
 Savvas A. Chatzichristofis Yiannis S. Boutalis Konstantinos Zagoris  
 Department of Electrical & Computer Engineering,  
 Democritus University of Thrace, Xanthi, Greece  
 {lazatsoc, athakapo, nikodour, schatzic, ybout, kzagoris}@ee.duth.gr

**Abstract**—This paper tackles the problem of the user’s incapability to describe exactly the image that he seeks by introducing an innovative image search engine called TsoKaDo. Until now the traditional web image search was based only on the comparison between metadata of the webpage and the user’s textual description. In the method proposed, images from various search engines are classified based on visual content and new tags are proposed to the user. Recursively, the results get closer to the user’s desire. The aim of this paper is to present a new way of searching, especially in case with less query generality, giving greater weight in visual content rather than in metadata.

**Keywords**—Image Retrieval; Metadata; Image Annotation; Query Recommendation Systems.

## I. INTRODUCTION

In the last few years the idea of Computer-Human Interaction is evolving continuously with fast steps. It is believed that this is the most crucial and promising direction to research for the technological future of humanity. Nowadays, there are many applications belonging to Computer-Human Interaction systems. A great example of that is the web search engines on the Internet and more specific the image search engines. Images are a very important part of our daily life. Google, Bing, Ask or Flickr, which are visited by millions of people follow a web image searching procedure which is based on the keywords of the user and the metadata of the images. Metadata can be keywords, tags or any other information from the web page that the image belongs to. But, there are some problems with that technique. Sometimes, images are often available without any metadata. In other cases, the annotation of the images is not correct and does not correspond to the actual description of the image (noisy annotations). Furthermore, the disadvantage of using textual query (keyword) for the image search is that the user is not always fully capable of describing his exact wish. As it is commonly said, ‘one image is consisting of a thousand words’, so it is impossible for the seeker to describe the image exactly as it is. As a result, the images proposed to him by the web search engines are often not relevant and far from his desires.

On the other hand, there could be a method that is based only on low level features of the image without using any textual information. Content Based Image Retrieval (CBIR) is based on the visual content of the image, e.g., color, texture, shape or information from local patches. CBIR is defined as any technology that in principle helps to organize digital image archives by their visual content [1]. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR. But, the ‘weak spot’ of CBIR is that it seems to be notoriously noisy for image queries of low generality [2]. If the query image happens to have a low generality, early rank positions may be dominated by spurious results. As a result, the simultaneous employment of CBIR techniques and metadata were found to be significantly more effective and reduce the communication gap that exists between the Humans and the Computers more than the text-only and image-only baseline [3].

In this paper, we propose a new query expansion recommendation system which tries to reduce the semantic gap between ‘What’ the user wants to find and ‘How’ he describes it. Query expansion is the process of reformulating a seed query to improve retrieval performance in information retrieval operations. The proposed system initially uses several parsers to take images from Google, Bing and Ask image search engines for a given textual query (keyword). More details about parsers are given in Section 2. Then, utilizing Color and Edge Directivity Descriptor (CEDD) [4] the visual content of these images are described. More details about the description of the visual content are given in Section 3. Next, the well-known K-means classifier, separates the image descriptors in a preset or in a dynamically calculated number of clusters. In order to calculate the number of clusters needed for each image set, a Self Growing and Self Organized Neural Gas Network (SGONG) is employed. More details about SGONG are given in Section 4. The images, whose descriptors have the minimum distance from the center of each class, are considered to describe better the subject of the current class. Consequently, those descriptors are used to retrieve the top-K, visually similar, images from

a collection of ones parsed from Flickr, using the same keyword with the other three search engines. The manually annotated tags of these top-K images are retrieved and, subsequently, using Wu and Palmer [5] method, we are calculating the semantic similarity between these tags and classify them into  $C$  classes. This process is illustrated in Section 5.

Depending on the number that each tag appears, a tag cloud is constructed. Those tags will be used afterwards to improve the query and the search process. The entire procedure is described in details in Section 6 while early experimental results are drawn in Section 7. Finally, the conclusions are given in Section 8. A preliminary version of this paper has been presented in [6].

The ideal scenario looks like this: A user inputs the keyword 'Paris'. The application searches through the internet using the three popular search engines with this keyword and generates a pool of results. The CEDD of each image is computed and the results are classified into two classes. The images, whose distance from the classes center is the minimum, are used for the retrieval of top-K relevant images, among the ones parsed from Flickr, under the same keyword. From those images and after the semantic classification of their tags, two classes are constructed: one with the tags 'Paris Eiffel Tower' and another one with 'Paris Hilton'. By clicking on a tag the application repeats the whole process using the improved query. So, the user of our web search engine can communicate in more specific way with his computer, and finally through this improved interaction to find exactly what he wishes.

Several re-ranking by visual content methods has been seen before, but mostly in different setups than the one we consider or for different purposes [7], [8]. Some of them used external information, e.g., an external set of diversified images [9] or training data [10]. The authors in [11] proposed a similar to our approach tag recommendation system based on visual similarity. According to their approach, the main problem of today's image search is the ambiguous definition of tags given from users. They are trying to bypass this phenomenon by recommending, the closest to image, tags. This is achieved by extracting the low-level visual features of the image. In our paper, we take advantage of this optimization of tags given to images, so we can propose more effective queries to users and help them get better results according to their desires.

## II. PARSERS

The word 'parse' means to analyze an object specifically. Parsing refer to breaking up ordinary text. For example, search engines typically parse search phrases entered by users so that they can more accurately search for each word. Some programs can parse text documents and extract certain information like names or addresses.

In this occasion, parsers are used to search in the produced html (Hypertext Markup Language) page of each search engine (Google, Bing, Ask, flickr) to find the url of each image they return. Despite the fact that web search engines propose their Application Program Interfaces (APIs) for performance of text search, they do not have any APIs for image searching, except from Flickr. So, it was urgent need to generate parsers to bypass that serious obstacle.

## III. COLOR AND EDGE DIRECTIVITY DESCRIPTOR (CEDD)

The recently proposed Color and Edge Directivity Descriptor belongs to the family of Compact Composite Descriptors (CCDs)[12]. An important thing about CEDD is that it uses only 54 bytes per image for indexing them, rendering this descriptor suitable for use in large image databases. Also the results of CEDD are very effective so the descriptor is very suitable for this web application.

In the technical part, CEDD initially separates images into a preset number of blocks. Each image block is classified in to one, or more than one of the  $n = 6$  preset texture areas. Each texture area consists of  $m = 24$  sub-regions. Using 2 fuzzy systems, CEDD classifies the colors of the image blocks in a 24-color custom palette. Then texture extraction is achieved by a fuzzy version of five digital filters proposed by MPEG-7 Edge Histogram Descriptor, forming 6 texture areas. When CEDD is used to describe an image block, each section of the image goes through 2 units: 1) the color unit and 2) the texture unit.

The color unit classifies the image block into one of the 24 shades used by the system in a color area,  $m, m \in (0, 23)$ . The texture unit classifies the image into a texture area,  $n, n \in (0, 5)$ . The image block is classified in the bin  $n \times 24 + m$ . This process is repeated for all the image blocks. At the end, the histogram produced, is normalized within the region  $[0, 1]$  and quantized for binary representation in a three bits per bin quantization.

## IV. SELF-GROWING AND SELF-ORGANIZED NEURAL GAS NETWORK

The Self Growing and Self Organized Neural Gas Network (SGONG) [13] is an innovative neural classifier. This network was proposed in order to reduce the number of colors in a digital image. It collects the advantages of the Growing Neural Gas (GNG) and the Kohonen Self-Organized Feature Map (SOFM) neural classifiers. The main advantage of the SGONG network is that it controls the number of created neurons and their topology in an automatic way. This feature is very important for that web application because it provides a method to compute the number of the classes automatically.

There are also some other characteristics of this neural classifier:

- The dimensions of the input space and the output lattice are always identical.
- In order to determine the classes and to ensure fast convergence, it uses some criteria.
- Except for color components, the SGONG neural network can also be used if its entrance is other local spatial features.
- The color reduction results obtained are better than the other two techniques which it combines.
- For the training procedure, the Competitive Hebbian Rule (CHR) is used to dynamically create or remove the connections of neurons.

As it is mentioned above, the SGONG uses some criteria in order to determine the number of created neurons. At the end of each epoch, three criteria that modify the number of the output neurons and make the proposed neural network to become self-growing are applied. These criteria are applied in the following order:

- remove the inactive neurons,
- add new neurons,
- and finally, remove the non important neurons.

### V. WORDNET-BASED SEMANTIC SIMILARITY MEASUREMENT

To define semantic similarity we have to consider the calculation of the conceptual similarity between words that are not lexicography similar such as the word ‘car’ and the word ‘automobile’. These procedures is accomplished by comparing the results of their relationship with a third ontology (like ‘wheeled vehicle’ in our example). In general semantic similarity helps to detect duplicate (high scores) or complementary (medium scores) content. In bibliography there are several methods for this particular job. The four most important are [14]:

- the edge counting method
- the information content method
- the feature based method
- the hybrid method.

Wordnet version 3.0 (2006) [15] is a lexical database which is available online and provides a large repository of English lexical terms. Wordnet was designed to establish the connections between four type parts of Speech (POS) such as noun, verb, adjective and adverb. Those POS’s are grouped into synonyms sets called synsets which are the smallest units in Wordnet and represent terms or concepts. The synsets are also organized into ‘senses’ which basically represent different meanings of the same term.

To calculate the semantic similarity between two synsets, the path length measurement was used. In [16], the authors are using a similar to our approach for enchanting search privacy on the Internet, focusing on plausible deniability against search engine query-log. Path length uses hyponyms and hypernyms. The hypernym represents a certain set of

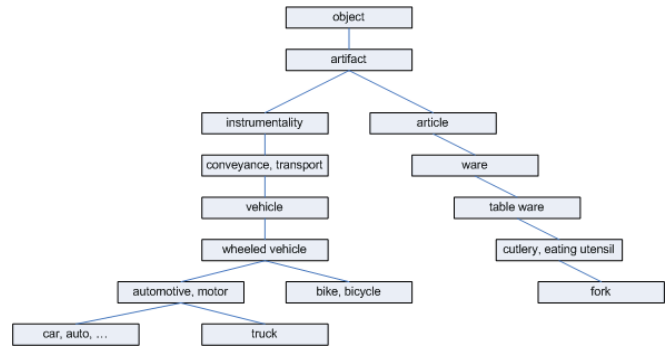


Figure 1. Example of the Hyponym Taxonomy in Wordnet Used For Path Length Similarity Measurement

discrete objects and the hyponym represents a smaller part of the hypernym. The taxonomy of the two synsets is treated as an undirected graph and measures the distance between them in Wordnet. In this graph, a shared parent of two synsets is known as a sub-sumer. The Least Common Sub-sumer (LCS) of two synsets is the sumer that does not have any children that are also the sub-sumer of two synsets. In this paper the method followed for this process is the Wu and Palmer similarity metric. This method measures the depth of the two synsets in the Wordnet taxonomy, and the depth of the least common sub-sumer (LCS) and combines these figures into a similarity score given below:

$$Sim = \frac{2 \times \text{depth(LCS)}}{\text{depth(Synset1)} + \text{depth(Synset2)}} \quad (1)$$

In contrary to the other dictionaries, Wordnet does not have any information about etymology, pronunciation and the forms of irregular verbs but only bounded information about the usage.

The actual lexicographical and semantic information is maintained in lexicographer files, which are then processed by a tool called grind to produce the distributed database. Both grind and the lexicographer files are freely available in a separate distribution, but modifying and maintaining the database requires expertise.

### VI. IMPLEMENTATION - METHOD OVERVIEW

All the technical characteristics described before are combined into an image web search engine called TsoKaDo.

At first, the user is asked to provide the search engine with an initial query, the number of images to be fetched from each search engine ( $M$ ) and the number of classes to be created ( $K$ ). Using, the user’s query, TsoKaDo fetches the top- $M$  results from the well-known search engines Google [17], Bing [18], Ask [19] and Flickr [20]. The difference between those search engines is that the Flickr store human imported tags for each image and this is TsoKaDo’s source of tags. A second difference between those is in their parsers as it mentioned in Section II.



In order the image’s visual information to be used, the Color and Edge Directivity Descriptor (CEDD) is extracted. The CEDD will produce a vector of 144 numbers that describes the color and the texture areas of the image. These vectors will be used from now on, and each consequent process will be taking place in the  $R^{144}$  space.

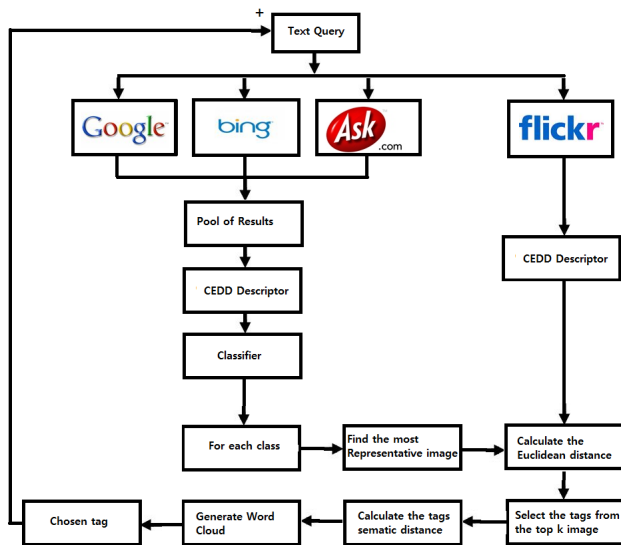


Figure 2. Steps Followed by Tsokado Image Web Search Engine.

At the next step, the pool of results of the first three search engines (Google, Ask, Bing), which is defined as:

$$P = M_{\text{Google}} \cup M_{\text{Bing}} \cup M_{\text{Ask}} \quad (2)$$

are classified using K-means algorithm or the Self-Growing and Self-Organized Neural Gas network (SGONG), depending on the users choice. If the user has selected an explicit number of classes ( $K = 2, 3, 4, 5$ ) only the K-means algorithm is used. Otherwise, if the option ‘Auto’ is selected, SGONG will propose the number of classes needed for reliable classification and the K-means will classify the images.

For each class, having its center available, TsoKaDo determines the image that has the least Euclidian distance from the class center. This image is considered to be the most representative image of the whole class and it is compared with each image returned from Flickr in the first step. The comparison is being performed by calculating the Euclidian distance and the tags of the image with the least one are fetched to describe the class.

Finally, in order to filter the tags collected, TsoKaDo uses Wordnet to calculate the semantic distance between them and stores them in a  $Y \times Y$  array, where  $Y$  the number of tags for the class. These are sorted according to their appearance frequency and then are proposed to the user by a tag cloud.

## VII. EXPERIMENTAL RESULTS

In order to show the functionality of TsoKaDo, some representative examples will be presented in this section.

For the first example, the keyword ‘Greece’ is chosen and various images about Greece are being retrieved from the search engines such as landscapes, islands and maps. These images are being classified into classes as it is shown in Figure 3.



Figure 3. Classification of Images Under the Keyword ‘Greece’.

Finally, new keywords are being proposed to the user for each class (see Figure 4). The tag cloud of class C, illustrated in Figure 4, contains the keywords ‘Parthenon’ and ‘Thessalonica’ and Class A tag cloud contain the terms ‘Athens’, ‘Cyclades’ and ‘Santorini’, which are all famous and popular destinations for vacations in Greece, also shown in the images of the class. In addition, class B contains many maps available on the web about Greece.



Figure 4. Tag clouds for Some Classes of Figure 3.

The user proceeds to the next step choosing a keyword to expand his search. In this example we choose to search more about ‘Santorini and the search engines response is shown in Figure 5 and the tag clouds in Figure 6, respectively.

As it is clearly shown, the new results of TsoKaDo are about the island Santorini only. In the class A there are some maps of the island and in the other classes some photos and landscapes. If the user wants to expand further his search, there is the keyword ‘stairs’ available and the final result is shown in Figure 7. To conclude with, this example demonstrates how effective the search can be extended with new keywords based on visual information.

This second example intends to demonstrate the advantages of using multiple search engines. In Figure 8 presents



Figure 5. Classification of images with query extended to 'Greece Santorini'.

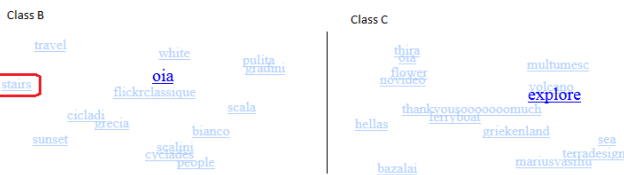


Figure 6. Tag clouds for Some Classes of Figure 5 - Tag: 'Santorini'.

the top results of the three such engines that TsoKaDo uses, given the keyword 'food'.

From the examples above, we conclude to the fact that TsoKaDo has two major advantages against conventional search engines:

To begin with, it combines the results of three search engines which very often return completely different results. It is known that every search engine stores a rank of web pages according to their importance based on some criteria.



Figure 7. Filtering 'Santorini' Results Using 'Stairs' Tag.

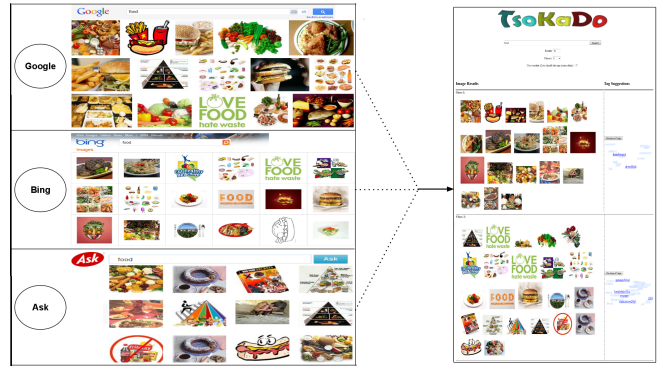


Figure 8. Comparison's results between TsoKaDo and the other search engines (for the same query 'food')

Combining three search engines gives the advantage of using three different ranks so it is even better than fetching more results from one engine regarding the diversity and quality of the images.

Furthermore, the most important advantage is that TsoKaDo manages to propose useful tags to the user. As it is mentioned earlier, the source of tags of this search engine is the Flickr which contains a huge database of images and photographs tagged with human imported keywords. This offers satisfying semantic recognition of objects, persons and locations that are depicted. The downside of using human imported tags from Flickr is that sometimes the keywords fetched are 'noisy'. As it is shown in Figure 9, besides the useful keywords there are many irrelevant terms. This can be improved at some point using Wordnet but this process may lose some useful data too. In Figure 9, it is shown how a tag cloud of a class with the query 'food is filtered using WordNet.

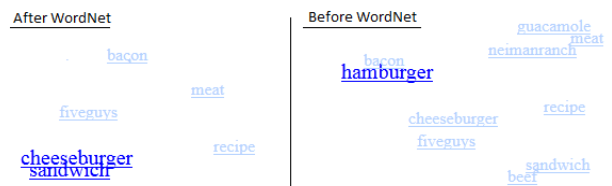


Figure 9. Filtering of tags using WordNet.

### VIII. CONCLUSIONS AND FUTURE WORK

In this paper, a new method in query expansion was proposed. With this method, users can expand their query with new keywords that are proposed to them relying on the extraction of low level visual features. This new image web search engine called TsoKaDo has a result of images, more corresponding to the users requirements.

Recently, Google introduces a new feature named 'sort by subject' [21], which at first may seem particularly close

to TsoKaDo's function. However, this is not true because Google does not use visual information of the image. Given that Google has access to almost every site in the web, it can determine keywords that in many important sites appear together. So it assumes they are relevant and proposes them to the user.

Although TsoKaDo offers something new in the field of image searching, it still has problems to deal with. At first TsoKaDo creates these classes because CEDD is based only in the color and the texture of the image; so, it cannot make an absolute semantic grouping of the images. In addition, the tag clouds may not be so relevant because the tags taken by Flickr, which are attached to every image by its users, are very noisy and don't always correspond to the content of the image. For example, in many cases, uploaders are using as tag for the images that the upload, details about the digital camera, that they use. This type of information does not describe the content of the image.

To bypass the current TsoKaDo problems, a wide range of future work can be expected. At first visual words might offer better semantic recognition of objects inside the images. Furthermore, the obstacle of 'noisy' tags can be overran by extending the search of relative images and keywords to more reliable sources such as Wikipedia. Wordnet can be, also, replaced by EuroWordnet, the Multilanguage version of Wordnet with various European languages for a better sorting of the tags. In addition, in order to measure the effectiveness of the proposed scheme a detailed case study will be performed. An on-line early version of TsoKaDo is available at [22].

#### REFERENCES

- [1] R. Datta, D. Joshi, J. Li, and J. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys (CSUR)*, vol. 40, no. 2, p. 5, 2008.
- [2] A. Arampatzis, K. Zagoris, and S. Chatzichristofis, "Dynamic two-stage image retrieval from large multimodal databases," *Advances in Information Retrieval*, pp. 326–337, 2011.
- [3] Arampatzis, A. and Zagoris, K. and Chatzichristofis, S., "Fusion vs. two-stage for multimodal retrieval," *Advances in Information Retrieval*, pp. 759–762, 2011.
- [4] S. Chatzichristofis and Y. Boutalis, "Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval," in *Proceedings of the 6th international conference on Computer vision systems*. Springer-Verlag, 2008, pp. 312–322.
- [5] Z. Wu and M. Palmer, "Verbs semantics and lexical selection," in *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 1994, pp. 133–138.
- [6] Lazaros T. Tsochatzidis, Athanasios C. Kapoutsis, Nikolaos I. Dourvas, S.A. Chatzichristofis, Yiannis S. Boutalis, "Query expansion based on visual image content," in *5th Panhellenic Scientific Conference for Undergraduate and Postgraduate Students in Computer Engineering, Informatics, related Technologies and Applications*, 2011.
- [7] K. Barthel, "Improved image retrieval using automatic image sorting and semi-automatic generation of image semantics," in *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on*. IEEE, 2008, pp. 227–230.
- [8] R. van Leuken, L. Garcia, X. Olivares, and R. van Zwol, "Visual diversification of image search results," in *Proceedings of the 18th international conference on World wide web*. ACM, 2009, pp. 341–350.
- [9] A. Popescu, P. Moellic, I. Kanellos, and R. Landais, "Lightweight web image reranking," in *Proceedings of the seventeen ACM international conference on Multimedia*. ACM, 2009, pp. 657–660.
- [10] T. Berber and A. Alpkocak, "Deu at imageclefmed 2009: Evaluating re-ranking and integrated retrieval systems," *CLEF Working Notes*, 2009.
- [11] M. Lux, A. Pitman, and O. Marques, "Can global visual features improve tag recommendation for image annotation?" *Future Internet*, vol. 2, no. 3, pp. 341–362, 2010.
- [12] Savvas A. Chatzichristofis, Yiannis S. Boutalis, *Compact Composite Descriptors for Content Based Image Retrieval: Basics, Concepts, Tools*. VDM Verlag Dr. Muller, 2011.
- [13] A. Atsalakis and N. Papamarkos, "Color reduction and estimation of the number of dominant colors by using a self-growing and self-organized neural gas," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 7, pp. 769–786, 2006.
- [14] G. Varelas, E. Voutsakis, P. Raftopoulou, E. Petrakis, and E. Miliotis, "Semantic similarity methods in wordnet and their application to information retrieval on the web," in *Proceedings of the 7th annual ACM international workshop on Web information and data management*. ACM, 2005, pp. 10–16.
- [15] G. Miller, "Wordnet: a lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [16] A. Arampatzis, P. Efraimidis, and G. Drosatos, "Enhancing deniability against query-logs," *Advances in Information Retrieval*, pp. 117–128, 2011.
- [17] [Online]. Available: <http://www.google.com>
- [18] [Online]. Available: <http://www.bing.com>
- [19] [Online]. Available: <http://www.ask.com>
- [20] [Online]. Available: <http://www.flickr.com>
- [21] [Online]. Available: <http://www.google.com/landing/imagesorting/>
- [22] [Online]. Available: <http://tsokado.nonrelevant.net/>

## Head Nod and Shake Gesture Interface for a Self-portrait Camera

Shaowei Chu

Department of Computer Science  
University of Tsukuba  
Tsukuba, Ibaraki, Japan  
chushaowei@iplab.cs.tsukuba.ac.jp

Jiro Tanaka

Department of Computer Science  
University of Tsukuba  
Tsukuba, Ibaraki, Japan  
jiro@cs.tsukuba.ac.jp

**Abstract** – Interactive interfaces and applications are a flourishing research area. In this paper, we introduce a head gesture interface for a digital camera shooting self-portrait pictures. Natural head nodding and shaking gestures can be recognized in real-time, using optical-flow motion tracking. A double head nod triggers the camera shutter to take shots. Continuous nodding or shaking triggers a zooming interface to zoom the user’s face in or out. To make the recognition robust, a safe zone analysis of the head region was conducted to quickly exclude any insignificant head motion, and thresholds of moving direction and length of head motion were selected in a preliminary set-up step. A finite state machine was used to recognize head gestures. Our results show that the proposed head gesture recognition method is a promising interface for a self-portrait camera.

**Keywords**-head gesture; self-portrait; human computer interaction; optical-flow; motion tracking.

### I. INTRODUCTION

Due to advances in digital cameras, including those in certain smart phones and foldable Liquid-Crystal Display (LCD) screens, taking self-portraits has become much easier. However, a user must physically touch the camera to change the frame or any settings, or in some cases, can use a remote control, but this can result in unnatural poses of the hands in pictures.



Figure 1. With hand gestures, a user can interact with a camera to trigger the shutter and take self-portraits.

We believe that one of the next major trends in the advancement of camera design will be making it more interactive, responsive, and accessible to the user. Applying

face, smile, and motion detection [10] functions to a camera is a good step, but does not fully satisfy users because of the modest degree of interaction. In our previous work [1], we proposed hand gestures for self-portrait photos, making the camera more interactive (Fig. 1), but this had limited success. Obviously, the system used in that study (and shown in the figure) is far too large to be incorporated into a portable digital camera; it requires a large display to provide a live view, where the user can see her/his gestures, visual tooltips, and a GUI. In practical applications, the screen of a camera will always be small, making it difficult or impossible to see such details.

Thus, we propose a head gesture interface, which unlike hand gestures, has no strict requirement for visual feedback and thus a small screen is acceptable (as shown in [3]), and it works well even with no feedback device (as shown in [6]). Head gestures can express clear meanings (e.g., a nod means yes and a shake means no), which are difficult to achieve using hand gestures. Moreover, when designing a zooming interface for self-portrait photos, it is difficult to develop a hand gesture-based interface, because the hands may extend outside the camera view when a zooming-in function is executed.

In this paper, we present and implement a head gesture interface, which triggers the camera shutter with a natural double nod and controls the zoom function with continuous nodding and shaking gestures (see Fig. 2).



Figure 2. Using the self-portrait application outdoors.

A Canon 60D digital camera was used, which provides hardware-supported face detection with a 30 FPS at  $1056 \times 704$  resolution video stream. The LCD screen is used as a front-facing screen to provide a live view to the user. For head gesture recognition, a safe zone analysis of the face

region is first conducted to quickly exclude large head motions. The features inside the face region are extracted only when the head motion is minor and is restricted to the safe zone and within a predefined period parameter (currently 500 ms). Then, after the features of the face region are selected, feature tracking is performed in each consecutive frame, and feature motion is recorded to compute the 2D head motion direction and length. Afterwards, user head motion data, in terms of direction and length, are collected through a user test step. Finally, based on the pattern analyzed from the user data, a Finite State Machine (FSM) is used to recognize the head gestures. In our implementation, head nods and shakes can be counted. The proposed interface was positively received in various user experiments.

The rest of this paper is organized as follows. Section 2 introduces related work on head gesture interfaces, Section 3 discusses our implementation, Section 4 introduces the interface design for self-portraits, Section 5 discusses preliminary user experiments and results, and Section 6 provides conclusions, a summary of the proposed interface, and possible future work.

## II. RELATED WORK

Vision-based gesture recognition is believed to be an effective technique for human-computer interaction, as presented in [11]. Some researchers have tried to apply hand gesture with digital cameras for taking self-portraits [1], and have had significant success. The benefits of such an approach are obvious: no additional devices or refitting of the camera are required. However, in particular cases, hand gestures may not be appropriate for self-portraits. For example, when zoomed in, the hands may extend outside the field of view. As a result, we explored a head gesture interface, because the face will always be within the camera view when taking a self-portrait. Furthermore, various studies of head gesture recognition [3][4] and head gesture interfaces [2][6] have shown that a head gesture interface is promising for self-portraits.

To date, there are two main approaches for recognizing head gestures. The Hidden Markov Models (HMM) method uses a pre-training process to collect user data, and then a pattern-recognition algorithm to distinguish between specific gestures [4]. A problem with this approach is that pre-training is required and there is a limited number of head gestures with two delayed digital outputs (nod or shake); in addition, few studies have achieved a recognition rate better than 85%. The other approach uses an FSM-based recognition technique [2][3] to explore temporal information on head motion in each video frame, and switches the states between head moving tendencies. Advantages of this method are that it involves less intense computing than HMM, and an adaptive threshold can be set in user experiments. We believe that a properly constructed FSM is the best option for this task.

In the present work, we provide a precise estimation of head motion direction and motion length in each frame from the video stream with about 50 features. In addition to template-matching of feature points [2], we apply an optical-

flow [7] measurement to tracking the motion of the points. Then we divide the tracking region into four sections (Up, Down, Left, and Right) using data collected from experiments. The FSM recognition is constructed based not only on timing and motion direction, but also the motion length. This implementation provides precise results requiring little computation.

We demonstrate that head gestures may be a promising interface for self-portraits. Our proposed method is just a beginning in this area of study.

## III. GESTURE RECOGNITION

Our implementation of a head gesture recognition algorithm must meet important requirements for real-time responsive applications: automatic initialization, sufficient sensitivity to recognize natural gestures, and an instantaneous and real-time response with feedback to the user. These requirements guided the development of our head gesture recognition scheme.

### A. Camera Device and Face Detection

Gesture recognition of head motion is based on face detection. In addition to implementing a face detection algorithm, we chose the Canon 60D camera, which offers SDK for developers, including embedded hardware support for face detection (Fig. 3).



Figure 3. Canon 60D camera with foldable LCD screen and face detection.

The camera runs at 30 FPS with  $1056 \times 704$  video image sequences for live view and has a foldable LCD screen.

### B. Safe Zone to Exclude Large Movements

After obtaining an image with a face region from the video stream, the image is processed to recognize head gestures. The first task is to define a safe zone that excludes large head motions that we assume are not nod or shake gestures. A safe zone, 40% larger than the face region, is initialized based on the face region in the image; then, in the following frames the system needs to check whether the face moved out of that safe zone. If so, then it is considered a large or fast motion, not a nod or shake gesture. The safe zone is then re-initialized and a new check begins. If not, then the safe zone remains as it is and the system waits for the next frame and face region to check again for head motion. When the face does not move out of the safe zone

within a specific period (currently, 500 ms), then we assume the head is still and it may be a good time to recognize head gestures. See Figure 4.

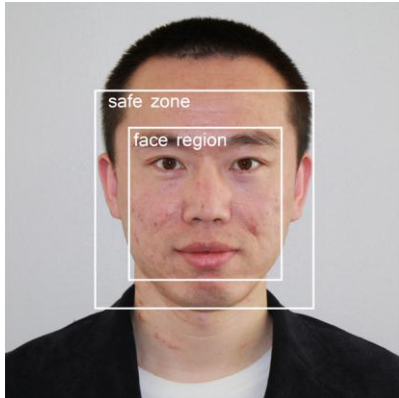


Figure 4. Face region and safe zone.

The safe zone is a key mechanism in the automatic initialization principle mentioned at the beginning of Section 3. Tracking, described in the next section, is reinitialized when the face moves out of the safe zone.

C. Feature Points Tracking and Motion Calculation

The features and their motion are calculated in consecutive images from the camera at 30 FPS, and the safe zone of the face is scaled to a 140 × 140 resolution size to perform feature extraction and tracking. There are two reasons for this. First, it identifies the face region size according to person and circumstantial context, such as user distance to the camera, helping to achieve a domain-independent model theory [4]. Second, it allows the calculation to be steady and fast for real-time application.

Within the face safe zone, we extract image features for tracking head motion. Several feature-extracting algorithms are known [8][9], but we chose a fast extracting method derived from the Hessian matrix, and selected the top 50 feature points as good features to track, as defined by Shi and Tomasi (S-T) features.

The Hessian-defined features rely on a matrix of the second-order derivatives ( $\partial^2x$ ,  $\partial^2y$ ,  $\partial x$ ,  $\partial y$ ) of image intensities. These are used because they are not sensitive to light. For each pixel point (x, y) in a second derivative image, the autocorrelation matrix over a small window around it is calculated, as follows:

$$M(x,y) = \begin{bmatrix} \sum_{-K \leq i,j \leq K} w_{i,j} I_x^2(x+i,y+j) & \sum_{-K \leq i,j \leq K} w_{i,j} I_x(x+i,y+j) I_y(x+i,y+j) \\ \sum_{-K \leq i,j \leq K} w_{i,j} I_x(x+i,y+j) I_y(x+i,y+j) & \sum_{-K \leq i,j \leq K} w_{i,j} I_y^2(x+i,y+j) \end{bmatrix}, \quad (1)$$

where  $w_{i,j}$  is a weighting term that can be defined as uniform or used to create a circular window around a pixel, and  $I$  is the intensity of a pixel. Good S-T features are found and

placed in the image where the autocorrelation matrix of the second derivatives has two large eigenvalues of the matrix.

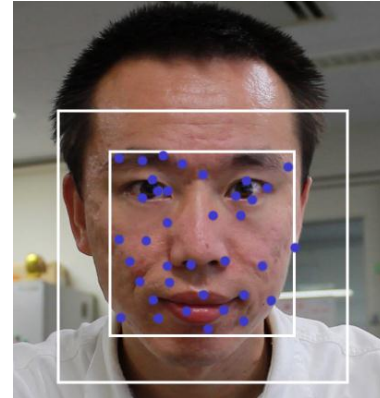


Figure 5. Numerous feature points are extracted.

After features are extracted (Fig. 5), a Lucas-Kanade [7] optical-flow measurement is taken to track each feature point's motion. The Lucas-Kanade method assumes that the displacement of the image contents between two nearby instant frames is small and approximately constant within the neighborhood of the point P under consideration. Thus, the optical flow equation can be assumed to hold for all pixels within a window centered at P. Moreover, by tracking the features with a pyramid layer of images, precise measurement of the velocity of feature motion can be achieved. This method functions well for tracking S-T features in the image. The optical-flow measurement results are feature points in the current frame's displacement from the previous frame. By calculating each feature's displacement, the motion length and direction of each feature point becomes clear.

Because of the error rate in optical-flow tracking, certain points will report a wrong result or be lost to tracking during head motion. Increasing the number of feature points is helpful to obtain reliable motion information. We calculate a set of feature points within the face region and mean values of length and direction as the main parameters in each frame.

It takes approximately 2.6 ms to extract the top 50 feature points and about 3.4 ms to track them using optical-flow measurements, which will suffice for real-time applications.

D. User Experiments

A user test step is applied to obtain head motion direction and length in frames while the user nods and shakes his/her head. The collected data are used to design the threshold of gesture recognition. The task is straightforward; users are asked to keep nodding or shaking during a period, and the moving direction degree and motion length in each frame with time is recorded.

A graduate student took part in this experiment to collect data. Measurements of the degree of motion begin at the top-right corner of the user's face as it appears in the display (Fig. 6). Figure 7 shows selected sequence data of nodding

and shaking gestures with motion degree (right vertical coordinate), length (left vertical coordinate), and timing (horizontal coordinate).

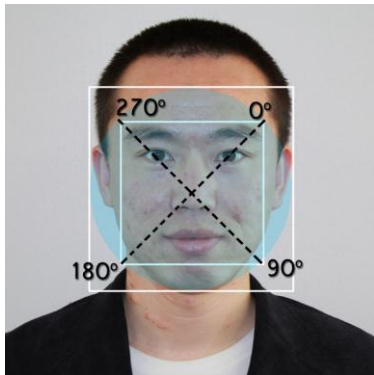


Figure 6. Motion direction angle coordinate.

Two important pieces of statistical information can be obtained from the data. First, when nodding (Fig. 7a), the motion directions are between 90° and 180° when tilting the head down and 270° and 360° when tilting the head up. In contrast, the motion directions in the shaking gesture (Fig. 7b) are smooth and steady along the 45° and 225° line. Second, the motion length data change periodically during the head gesture, but below a peak value of 6.

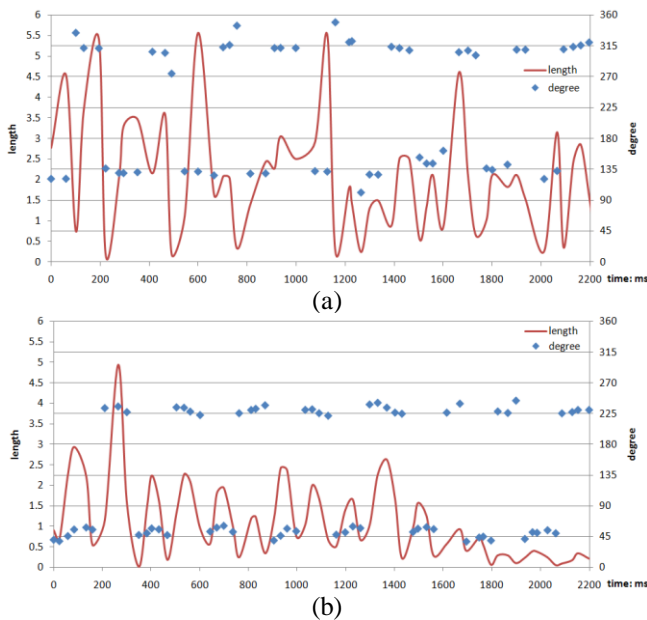


Figure 7. Nodding (a) and shaking (b) gesture data.

The time interval of up-down and left-right movement was 122.2 ms and 137.5 ms, respectively.

**E. Recognition Design**

Based on the data analyzed above, we can conclude that head shaking is a more steady motion than nodding. Thus,

we separated the motion regions for moving direction recognition into four regions (Fig. 8):

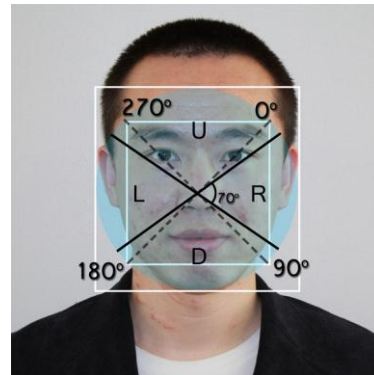


Figure 8. Motion direction region (U: Up, R: Right, D: Down, L: Left).

right, 10–80° (70° span); left, 190–260° (70° span); up, 260–360° and 0–10° (110° span); and down, 80–190° (110° span). The motion length value must be larger than 0.5 and less than 6.0 in the recognition procedure.

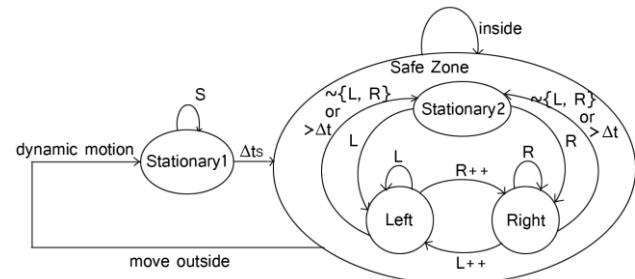


Figure 9. Finite State Machine for recognizing head shakes.

The timing-based FSM described in [2] was used to recognize the gestures. Figure 9 shows a transition chart for head shaking. It contains two main states: Stationary1, a motionless state, and Safe Zone, when the face is within the safe zone discussed above for a certain period of time. Inside the Safe Zone state, there are Stationary2, Left motion, and Right motion. The transition from Left to Right or reverse transition will add one factor to the shaking count.

**F. Performance and Implementation**

To increase the speed and efficiency of image processing, we used OpenCV [5] to implement the algorithm. The program was written in C++ with multi-threading. Image drawing was under Direct2D API support. The processing performance is shown in Table 1, based on an Intel Core 2 Quad CPU 2.5 GHz PC.

TABLE 1. IMAGE PROCESSING PERFORMANCE

Image processing	Process time (ms)
Resize face region	0.2
Extract features	2.6
Feature tracking	3.4
Gesture recognition	2.6

Note that feature extraction is performed on when initializing tracking; it is not executed in each frame. In addition, as shown in Table 1, in each frame with a 140 x 140 resolution safe zone, the process of tracking features requires the most computation time. Thus, one way to reduce the calculation time would be to further scale down the size of the safe zone. Finally, the gesture-recognition task, which calculates the motion of every feature point, depends on the number of points: tracking fewer points requires less time.

The Canon 60D camera is connected to the computer through a USB cable. Currently, the Canon 60D camera does not support programming to configure its foldable frontal screen. Thus, we developed an iPhone application to show the live view video and GUI. The iPhone application runs at about 13 FPS with a 360 x 240 resolution motion jpeg image through a WiFi connection.

#### IV. INTERFACE DESIGN FOR SELF-PORTRAITS

The original goal of this study was to design a natural head gesture interface for self-portrait photos. From the recognition procedure and the results described above, we designed an application based on counted nods and shakes. On the frontal screen, the user’s face region and safe zone are circled with lines. Graphical tooltips for the number of nods and shakes are drawn in the top area above the face region instantaneously when the user performs a gesture.

The double nod gesture (performing the nod gesture twice) triggers the camera shutter immediately. After the shutter is triggered, the self-timer is activated and a countdown (set at 5 s, because it takes 2–4 s to drive the mechanical lens and run the auto focus on the user’s face) is shown on the front-facing screen. At the same time, the camera adjusts the lens, focuses, and opens a flash when needed. The user can prepare her/his pose during this period and get ready for the self-portrait.

Continuous nodding or head shaking (when the gesture is performed three or more times) triggers the zoom function, where nodding zooms in and shaking zooms out. This function is performed with smooth gesture transitions.



Figure 10. Canon 60D camera with an iPhone as a front-facing screen.

The camera can be set on any steady object (e.g., by using a gorillapod [12]) or on a tripod indoors or outdoors. An iPhone was attached to the camera as a front-facing screen (Fig. 10).

#### V. DISCUSSION

Compared to our previous work on hand gestures, the head gesture interface is more suitable for providing a zoom interface. As mentioned in Section 1, it is not practical to use hand gestures for zooming. In addition, head gesture recognition is independent of user distance to the camera, which simplifies gesture recognition. Moreover, the implementation of the recognition algorithm does not vary depending on lighting conditions; it works well under any light conditions. Furthermore, using a frontal screen, the system is portable and can be taken outdoors.

While detailed user experiments have not yet been performed, informal user feedback from students at our university has been very positive. After a very brief introduction to the system, users were free to explore the system on their own. Feedback from students not specializing in computer science was more positive; they found the gesture-based manipulations to be intuitive and understandable, with descriptions such as “accessible” to describe the overall system and the general idea. The nod/shake counts shown above the face region of the display made it simple for users to understand, and they became familiar with the routine of recognition based on the visual tooltips.

The distance to the camera was an issue raised by one person; although our implementation is completely distance-independent, when a user is a significant distance from the camera, it obviously becomes difficult to see the preview on the frontal screen clearly. However, when indoors, the system can be plugged into a larger display, as we did in [1].

The automatic face tracking and zooming-in to the face region may not satisfy a user who wishes to have full-body pictures. In such a case, a pan and tilt platform could be used, and the zooming-in function could zoom in on the center of the image rather than on the face region.

The issue of taking profile pictures was also raised. This can be done by facing the camera and performing a double nod to trigger the shutter, then turning into the profile position and waiting for the picture to be taken. Another possibility would be to modify the system so that it can estimate the head pose based on eye location [3], and the adjust the thresholds of motion direction for profile nod and shake gestures.

For pictures of multiple users, a face recognition technique could be applied to a main user only, and functions could be triggered based only on that person’s gestures.

One participant suggested that our interface could be useful for handicapped persons, who could take hands-free self-portraits.

#### VI. CONCLUSIONS AND FUTURE WORK

We presented a head gesture interface for self-portraits. The shutter is activated by double nods, and zoom functions



are performed by continuous nods and shakes. In the gesture recognition procedure, a safe zone is used to exclude large, irrelevant motions. Numerous feature points are extracted and tracked based on optical-flow measurements, and FSM is used to recognize head gestures. Our system runs in real-time, counting nods and shakes. We performed trial runs and discussed the primary results, which showed that this method is a promising interface for self-portraits.

In the near future, flexible and context-based interfaces will be integrated into the system to support more configuration functions of the camera. In addition, more user experiments will be conducted to refine the system.

#### REFERENCES

- [1] S. Chu and J. Tanaka, "Hand gesture for taking self portrait," HCI International 2011, Human-Computer Interaction, Part II, LNCS 6762, pp. 238-247, Orlando, FL, USA, July 9-14, 2011.
- [2] J. Davis and S. Vaks, "A perceptual user interface for recognizing head gesture acknowledgements," Proceedings of the 2001 workshop on Perceptive user interfaces, pp. 1-7, Orlando, FL, USA, 2001, doi: <http://doi.acm.org/10.1145/971478.971504>.
- [3] R. Li, C. Taskiran and M. Danielsen, "Head pose tracking and gesture detection using block motion vectors on mobile devices," Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology, pp. 572-575, Singapore, September 10-12, 2007, doi: <http://doi.acm.org/10.1145/1378063.1378157>.
- [4] J. Lee and S. Marsella, "Learning a model of speaker head nods using gesture corpora," Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems, vol. 1, pp. 289-296, Budapest, Hungary, July 9-14, 2009.
- [5] Open Source Computer Vision Library (OpenCV): Last visited on November 22nd, 2011. <http://opencv.willowgarage.com/wiki/>
- [6] I. Yoda, K. Sakaue, and T. Inoue, "Development of head gesture interface for electric wheelchair," Proceedings of the 1st international convention on Rehabilitation engineering and assistive technology: in conjunction with 1st Tan Tock Seng Hospital Neurorehabilitation Meeting, pp. 77-80, Singapore, 2007, doi: <http://doi.acm.org/10.1145/1328491.1328511>.
- [7] J. Bodily, B. Nelson, Z. Wei, D. Lee, and J. Chase, "A Comparison Study on Implementing Optical Flow and Digital Communications on FPGAs and GPUs" ACM Transactions on Reconfigurable Technology and Systems, vol. 3, Issue 2, pp. 6:1-6:22, 2010, doi: <http://doi.acm.org/10.1145/1754386.1754387>.
- [8] L. Trujillo and G. Olague, "Automated design of image operators that detect interest points" Journal of Evolutionary Computation, vol. 16, Issue 4, pp. 483-507, 2008, doi: <http://dx.doi.org/10.1162/evco.2008.16.4.483>.
- [9] H. Bay, T. Tuytelaars, and L. Gool, "SURF: speeded up robust features," Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346-359, 2008.
- [10] Casio TRYX camera: Last visited on November 22nd, 2011. [http://exilim.casio.com/digital\\_cameras/TRYX/TRYX](http://exilim.casio.com/digital_cameras/TRYX/TRYX).
- [11] M. Pranav and M. Pattie, "SixthSense: a wearable gestural interface," ACM SIGGRAPH ASIA 2009 Sketches, pp. 11:1-11:1, 2009, doi: <http://doi.acm.org/10.1145/1667146.1667160>.
- [12] Joby Gorillapod: Last visited on November 22nd, 2011. <http://joby.com/gorillapod>.

## Towards 3D Data Environments using Multi-Touch Screens

Francisco Ortega, Naphtali Rishe  
*School of Computing and Information Sciences*  
*Florida International University*  
*Miami, FL, USA*  
*Forte007@fiu.edu, NDR@acm.org*

Armando Barreto, Malek Adjouadi  
*Electrical and Computer Engineering Department*  
*Florida International University*  
*Miami, FL, USA*  
*BarretoA@fiu.edu, Adjouadi@fiu.edu*

**Abstract**—The increase in availability of multi-touch devices has motivated us to consider interaction approaches outside the limitations associated with the use of a mouse. The problem that we try to solve is how to interact in a 3D world using a 2D surface multi-touch display. Before showing our proposed solution, we briefly review previous work in related fields that provided a framework for the development of our approach. Finally, we propose a set of multi-touch gestures and outline an experiment design for the evaluation of these forms of interaction.

**Keywords**-Multi-Touch; 3D Interaction ;3D Navigation; Gestures.

### I. INTRODUCTION

This paper presents the initial development of our approach to work with 3D interaction worlds using a multi-touch display. We introduce our emerging methods in the context of important previous work, resulting in our proposed translation and rotation gestures for multi-touch interaction with 3D worlds. In this same process, we try to answer two important questions and provide an evaluation path to be implemented in the near future.

3D navigation and manipulation are not new problems in Human-Computer Interaction (e.g., [1], [2]) as will be shown in our brief review of previous work. However, with the availability of multi-touch devices such as the iPad, iPhone and desktop multi-touch monitors (e.g., 3M M2256PW 22" Multi-Touch Monitor) new concepts have developed in order to help the transition to a post-Windows-Icon-Menu-Pointer (WIMP) era. This gives rise to important questions such as: (1) What is the most appropriate mapping between the 2D interface surface and the 3D world? and (2) Can previous techniques used with other devices (e.g., joystick, keyboard and mouse) be used in 3D navigation?

To begin answering these questions, we first endeavor to understand touch interactions and previous multi-touch work. We believe that all those aspects create a foundation that is necessary for the development of a sound post-WIMP framework [3]. After the related work section, we cover our proposed solution, discussion and future work.

### II. BACKGROUND

#### A. Understanding Touch Interactions

A common option for multi-touch interaction is to use the set of points corresponding to  $n$  touches in a direct manner. However, to achieve a more natural interaction between the screen and the user, studies like [4]–[8] provide a different take on how touch information can be used. For example, [4] studies finger orientation for oblique touches which gives additional information without having extra sensors (e.g., left/right hand detection.) In another example, Benko and Wilson [8] study dual finger interactions (e.g, dual finger selection, dual finger slider, etc.) Additional work dealing with contact shape and physics can be found in [6], [7]. In a very comprehensive review of finger input properties for multi-touch displays [5] provides suggestions that have been used already in [4].

One aspect that is important to have in mind is whether to keep rotations, translations and scaling separate [9] or combined [10]. If the latter is chosen, the user's ability to perform the operations separately may become a problem [9].

One very important point found in [11], [12] is that one-hand techniques are better for integral tasks (e.g., rotation) and two hands perform better with separable tasks. For our particular work, one can think of using one hand to perform common rotations and translations, and using two hands when special rotations need to be performed, utilizing the second hand to indicate a different point of reference.

#### B. Multi-Touch Techniques

We believe that all of the related work dealing with multi-touch, regardless whether it was designed for manipulation of objects or navigation of 3D graphical scenes, can contribute to the set of unifying ideas that serves as the basis for our approach.

Some of the work in 3D interactions has been specific for multi-touch, which is our focus as well. In [10], Hancock et al. provide algorithms for one, two and three-touches. This allows the user to have direct simultaneous rotation and translation. The values that are obtained from initial touches  $T_1$ ,  $T_2$  and  $T_3$  and final touches  $T'_1$ ,  $T'_2$  and  $T'_3$  are

$\Delta$ yaw,  $\Delta$ roll and  $\Delta$ pitch, which are enough to perform the rotation in all three axes and  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  to perform the translation. A key part of their study showed that users prefer gestures that involve more simultaneous touches (except for translations). Using gestures involving three touches was always better for planar and spatial rotations [10].

A different approach is presented in RNT (Rotate 'N Translate) [13], which allows planar objects to be rotated and translated using opposing currents. This particular algorithm is useful for planar objects and it has been used by 3D interaction methods (e.g., [14]).

A problem that occurs when dealing with any type of 3D interaction in multi-touch displays, is the one of spatial separability [9]. To address this problem, in [9], the authors proposed different techniques that will allow the user to perform the correct combination of transformations (e.g., scaling, translation + rotation, scaling + rotation + translation, etc.). The two methods that were most successful were Magnitude Filtering and First Touch Gesture Matching. Magnitude Filtering works similarly to snap-and-go [9]. This method has some differences from normal snapping techniques because it does not snap to pre-selected values or objects. In addition, the authors introduce a catch-up zone allowing “continuous transition between the snap zone and the unconstrained zone.” [9]. The latter method, First Touch Gesture Matching, works by minimizing “the mean root square difference between the actual motion and a motion generated with a manipulation subset of each model” [9]. To select the most appropriate model, each prospective model creates two outputs, best-fit error and magnitude of the appropriate transformation. These outputs are given to an algorithm that decides which models to apply.

### III. PROPOSED SOLUTIONS

#### A. Set up

1) *Camera:* To test our work, we use OpenGL and perform the visualization through a virtual camera developed in [15] and described by [16], as shown in Figure 1. One can see that the UP vector indicates which way is up, the EYE (or ORIGIN) vector indicates the position of the camera and the AT (or FORWARD) vector indicates the direction in which the camera is pointing.

2) *Multi-Touch:* We are using Windows 7 multi-touch technology [17] to test our proposed solutions using a 3M M2256PW Multi-touch Display. Windows 7 provides two ways of using the touches from the device. The first one is gesture-based, identifying simple pre-defined gestures and the second is the raw touch mode which provides the ability to develop any type of interaction. We chose the latter because our end goal is to create custom user interactions and for that we preferred to work at the lowest level possible that is available to us. Each touch has a unique identification (ID) that is given at the moment of TOUCHDOWN, to be used during the TOUCHMOVE, and to end when TOUCHUP has

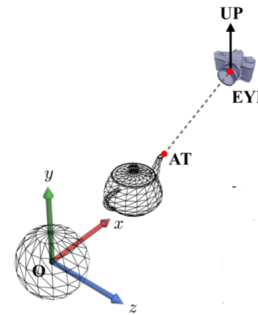


Figure 1: Camera View [16]

been activated. The ID gives us a very nice way to keep track of a trace, which is defined as the path of the touch from TOUCHDOWN to TOUCHUP.

3) *Visual Display:* As a test case, we have created a world with 64 by 64 by 64 spheres, in a cubic arrangement, drawn in perspective mode, where each sphere has a different color. This allows the user to test the 3D navigation provided by our gestures while having a visual feedback. It is important to note that we colored the spheres in an ordered fashion using the lowest RGB values in one corner and the highest values in the opposite corner of the cube of spheres.

#### B. Gestures

We have decided to develop separate gestures for translation and rotation to understand what combinations are more efficient for the user. This means that when rotating, the user will be rotating by a specific axis and translations will be performed using two axes. We also decided to provide simple gestures for our initial design to see the interaction of the users. Once we have collected more data about the interaction, we can create more complex gestures, if needed. Before we discuss translation and rotation techniques, we will clarify the notations used. Lowercase x, y and z refer to 3D axes, Xscreen and Yscreen refer to 2D axes in the screen (using OpenGL view mode,) and uppercase X,Y and Z refer to a point in the x, y, z axes.

1) *Translation:* In order to translate the camera, we decided to combine the x & y axes and leave z by itself. The algorithm for the y axis is similar to Algorithm 1, replacing the variable X with the variable Y. The algorithm for the z axis is similar to the y axis with the exception that it uses 3 touches. In general, the user can perform simultaneous translation by the x and y axes using one finger or translate by the z axis using three fingers. All the movements can be executed with a single hand.

2) *Rotations:* To address rotations, we have to think of rotation about x, y and z independently, given that is our belief that separating the rotation will demand a lower cognitive load from the user. This expectation is also supported by [9], [11], [12]. In addition, all the rotations are designed to use only one hand, which is preferable, as demonstrated

---

**Algorithm 1** Translation over X
 

---

**Require:** TouchCount = 1  
**if**  $Point.X < PrevPoint.X$  **then**  
     MoveRight(delta)  
**else if**  $Point.X \geq PrevPoint.X$  **then**  
     MoveRight(-delta)  
**end if**

---

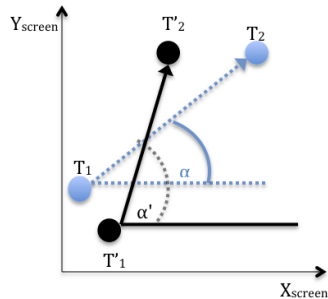


Figure 2: Rotation about the 3D x axis (see Figure 1)

in [9]. To keep the constraint of using only one hand, the algorithm checks that the touches are within a cluster.

The gesture for rotation about x, as shown in Figure 2, merits to be described in more detail because the other two rotations about y and z use very similar algorithms to those already described for the translations. The only difference is that y and z rotations require two touches each. The gesture for rotation about x begins with  $T_1$  and  $T_2$ , which form an angle  $\alpha$  with the horizontal axis. The user's final state is represented with  $T'_1$  and  $T'_2$ , forming an  $\alpha'$  angle with the horizontal. Then, the difference between  $\alpha'$  and  $\alpha$  gives the rotation angle to be applied, about x.

#### IV. DISCUSSION

##### A. Gestures

We believe that the set of gestures that we are proposing based on the literature reviewed in the background section and our own preliminary testing will give a starting point to find the most natural gestures to interact in a 3D world using a multi-touch device. Even after finding the most natural gestures for 3D navigation, one will have to compare with other devices such as the ones found in Bowman et al. [18]. As we will outline in the next section, we suggest to make the comparison with a 3D mouse, such as the one shown in Figure 3.

The first question asked in the introduction was: What is the most appropriate mapping between the 2D interface surface and the 3D world? We have proposed a simple solution to the problem, through the set of gestures described above. Defining and implementing the most natural mapping between this 2D multi-touch interface and the 3D world may still require additional work, but the concepts advanced in



Figure 3: 3D Mouse

this paper may provide an interesting direction towards the solution of this ongoing challenge.

The other question asked in the introduction was: Can previous techniques used with other devices (e.g., Joystick, keyboard and mouse) be used in 3D navigation? We propose that the answer is yes. We can build upon existing work that was developed for the mouse or other interfaces, adapting it for use with multi-touch displays whenever possible. An example of this is The Virtual Sphere [2]. We could take The Virtual Sphere and create a similar device for use with multiple fingers to allow a pure 3D rotation and translation, even emulating the 3D mouse shown in Figure 3. However, those considerations would be outside the scope of this paper.

In general, we find that multi-touch displays can work efficiently for achieving a more natural user 3D interaction and 3D manipulation.

##### B. Proposed Evaluation Technique

The considerations presented above inform our current process of planning the experimental protocol and data analysis methods we will use for evaluating our approach. To answer our research questions and test the proposed gestures, we will recruit at least 30 subjects from the college student population at our university. The reason for our choice of target population is that we believe that all students will have a good grasp of basic mouse interaction, which will facilitate the completion of the experimental tasks by the subjects.

The actual experiment, after allowing the user to become familiarized with the interface, will consist of a set of tasks to test translation and rotation gestures (independently) using our 3M 22" Multi Touch Monitor (Model M2256PW) and the 3D mouse made by 3DConnexion, which is shown in Figure 3. For each of the tasks, we will measure the time of execution to complete the task, and the accuracy of the movement. For the completion time, we will use an external game controller to start and stop the time, and for the accuracy of the movement, we will automatically record the initial and final positions. In addition to the automated recording of performance data, we will ask the subjects to complete a short usability questionnaire [19].

#### V. CONCLUSION AND FUTURE WORK

Recently, multi-touch displays have become more widely available and more affordable. Accordingly, the search for

protocols that will simplify the use of these devices for interaction in 3D data environments has increased in importance. In this paper we have outlined some of the most valuable previous contributions to this area of research, highlighting some of the key past developments that have emerged in the 3D-interaction community. This review of pertinent literature provides a context for the presentation of the core elements of the solution we propose for the interaction in 3D environments through a multi-touch display.

Specifically, we proposed a set of multi-touch gestures that can be used to command translations and rotations in 3 axes, within a 3D environment. Our proposed solution has been implemented using a 3M M2256PW 22" Multi-Touch Monitor as the interaction device. This paper explained the proposed gestures and provided pseudo-code segments that indicate how these gestures are to be captured using the information provided by the device. In our definition of the proposed multi-touch gesture set we have established independent gestures for each type of translation and also for each type of rotation. We decided to proceed in this way so that we can study how users prefer to combine or concatenate these elementary gestures.

The next step in the development of our approach is to evaluate its efficiency in a comparative study involving other 3D interaction mechanisms, such as a 3D mouse. The ongoing process of planning the experiments for evaluation takes into account the nature of the devices and general principles of design of experiments, in an effort to minimize the presence of confounding effects, such as subject fatigue, etc. Our experiments may lead us to define alternative gestures to allow more innovative means of interaction.

#### ACKNOWLEDGMENTS

This work was sponsored by NSF grants HRD-0833093, and CNS-0959985. Mr. Francisco Ortega is the recipient of a GAANN fellowship, from the US Department of Education, at Florida International University.

#### REFERENCES

- [1] G. Nielson and D. Olsen Jr, "Direct manipulation techniques for 3D objects using 2D locator devices," *Proceedings of the 1986 workshop on Interactive 3D graphics*, pp. 175–182, 1987.
- [2] M. Chen, S. Mountford, and A. Sellen, "A study in interactive 3-D rotation using 2-D control devices," *ACM SIGGRAPH Computer Graphics*, vol. 22, no. 4, p. 129, 1988.
- [3] R. Jacob, A. Girouard, L. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum, "Reality-based interaction: a framework for post-WIMP interfaces," *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems (CHI '08)*, pp. 201–210, 2008.
- [4] F. Wang, X. Cao, X. Ren, and P. Irani, "Detecting and leveraging finger orientation for interaction with direct-touch surfaces," *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pp. 23–32, 2009.
- [5] F. Wang and X. Ren, *Empirical evaluation for finger input properties in multi-touch interaction*. ACM, Apr. 2009.
- [6] A. Wilson, S. Izadi, O. Hilliges, A. Garcia-Mendoza, and D. Kirk, "Bringing physics to the surface," *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pp. 67–76, 2008.
- [7] X. Cao, A. Wilson, R. Balakrishnan, K. Hinckley, and S. Hudson, "ShapeTouch: Leveraging contact shape on interactive surfaces," *Horizontal Interactive Human Computer Systems, 2008. TABLETOP 2008. 3rd IEEE International Workshop on*, pp. 129–136, 2008.
- [8] H. Benko, A. Wilson, and P. Baudisch, "Precise selection techniques for multi-touch screens," in *In Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI '06)*, 2006, pp. 1263–1272.
- [9] M. A. Nacenta, P. Baudisch, H. Benko, and A. Wilson, "Separability of spatial manipulations in multi-touch interfaces," in *GI '09: Proceedings of Graphics Interface 2009*. Canadian Information Processing Society, May 2009.
- [10] M. Hancock, S. Carpendale, and A. Cockburn, "Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques," *Proceedings of the SIGCHI conference on Human Factors in computing systems*, p. 1156, 2007.
- [11] K. Kin, M. Agrawala, and T. DeRose, *Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation*. Canadian Information Processing Society, May 2009.
- [12] T. Moscovich and J. Hughes, "Indirect mappings of multi-touch input using one and two hands," *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems (CHI '08)*, pp. 1275–1284, 2008.
- [13] R. Kruger, S. Carpendale, S. Scott, and A. Tang, "Fluid integration of rotation and translation," *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 601–610, 2005.
- [14] J. Reisman, P. Davidson, and J. Han, "A screen-space formulation for 2D and 3D direct manipulation," *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pp. 69–78, 2009.
- [15] R. S. Wright, N. Haemel, G. Sellers, and B. Lipchak, *OpenGL SuperBible*, ser. Comprehensive Tutorial and Reference. Addison-Wesley Professional, Jul. 2010.
- [16] J. Han and J. Kim, *3D Graphics for Game Programming*. Chapman & Hall, Feb. 2011.
- [17] Y. Kiriarty, L. Moroney, S. Goldshtein, and A. Fliess, *Introducing Windows 7 for Developers*. Microsoft Pr, Sep. 2009.
- [18] D. A Bowman, "3D user interfaces: theory and practice," p. 478, Jan. 2005.
- [19] J. Lazar, D. Jinjuan Heidi Feng, and D. Harry Hochheiser, *Research Methods in Human-Computer Interaction*, Jan. 2010.

# User Interface for Trust Decision Making in Inter-Enterprise Collaborations

Puneet Kaur, Sini Ruohomaa and Lea Kutvonen

*Department of Computer Science*

*University of Helsinki*

*Helsinki, Finland*

*Email: puneet.kaur@cs.helsinki.fi, sini.ruohomaa@cs.helsinki.fi, lea.kutvonen@cs.helsinki.fi*

**Abstract**—Trust decisions on inter-enterprise collaborations involve a trustor’s subjective evaluation of its willingness to participate in the specific collaboration, given the risks and incentives involved. In earlier work, we have built support on automating routine trust decisions based on a combination of risk, reputation and incentive information. Non-routine cases must be dealt with by human users, who require access to supporting information for their decisions; further, their needs differ somewhat from the needs of automation tools. This paper presents work in progress to provide a usable user interface for manual trust decisions on inter-enterprise collaborations in situations where automated decisions cannot be made. We have implemented a trust decision expert tool and are in the process of evaluating it and incorporating it into a broader collaboration management toolset.

**Keywords**-trust decisions; reputation; risk; inter-enterprise collaboration; expert tool

## I. INTRODUCTION

The emergence of technology support ranging from service-oriented architecture and Web Services to cloud infrastructures are paving the way for semi-automated and low-cost setup and management of inter-enterprise collaborations. An inter-enterprise collaboration involves a network of autonomous enterprises working towards a shared goal, e.g., to provide a composed service to end users. An online travel agency, for example, can compose travel packages for its customers by utilizing a set of services provided by its partners for payment handling, booking flights, hotel itinerary, car rental and other location-specific arrangements, where each partial service is provided by a separate autonomous enterprise.

Inter-enterprise collaborations are particularly useful for small and medium-sized enterprises, which hold expertise in their own domain but have limited resources. By collaborating with other enterprises, they can attain a competitive edge in fields outside their individual scope, and also join forces to expand their business into fields dominated by large enterprises [1], [2]. Large enterprises can apply the same methods to organize their production life-cycles in-house, or to experiment on new service concepts together with external collaborators.

The success of inter-enterprise collaborations relies on a flexible infrastructure that reduces the cost of setting up and managing the collaborations, so that individual enterprises

do not need to solve issues of interoperability, collaboration coordination and trust management using costly ad hoc solutions. The Pilarcos open service ecosystem we have proposed in earlier work [2] provides infrastructure services for, e.g., finding potential partners and ensuring service interoperability, collaboration management and semi-automated trust decisions [1]. In this paper, we focus on the trust management support specifically.

Inter-enterprise collaboration depends on trust, as the autonomy of partners causes uncertainty and risk that must be found acceptable for the collaboration to proceed. The distributed infrastructure services in Pilarcos allow the enterprises to make local, private trust decisions on whether they want to join or continue in an inter-enterprise collaboration, subjectively analyzing the risks and incentives involved in the endeavour [3]. To ensure a combination of efficiency and a swift reaction to any major changes in the risk estimations, routine decisions are automated, following local policies [1].

Automated trust decisions can only be relied on in routine cases: human intervention is required for making trust decisions in situations where the risk or incentives are particularly high, or the information available for supporting the decisions is insufficient. In earlier work, we have set the basis for how to identify these special cases in policy, including support for, e.g., measuring the amount and quality of the input information [1], [4], and investigating the information needs of human users in this context [5]. Our goal in this paper is to provide a usable user interface for human intervention in semi-automated trust decisions on inter-enterprise collaborations.

The rest of the paper is organized as follows: Section II discusses related work in human trust decision making, and summarizes our earlier findings on the topic. Section III provides an overview of the Pilarcos trust management system and compares it to related work. Section IV presents the implemented expert tool user interface. Section V presents initial results of user evaluations, and Section VI concludes the paper.

## II. TRUST DECISIONS ON INTER-ENTERPRISE COLLABORATIONS

Trust decisions measure a subjective willingness of a trustor to perform a given action with a given trustee,

considering the risks and incentives involved [3]. In the context of our work, both the trustors and trustees are business services; this level of abstraction reflects the fact that two services even within the same enterprise may hold different information, have a different effect on assets, and be governed by different policies. When a trust decision is delegated to a human user, therefore, the interventions are made on behalf of a specific service.

Some of the expected risks and gains can be estimated based on past behaviour, i.e. the reputation of the trustee, while other incentives are created by the business importance of the activity itself, such as a need to fulfil existing contracts, or a desire to try out a new way of making business or a new set of partners.

A large body of existing work on reputation systems has focused on electronic markets, aiming to support either human or automated decisions specifically. Work in this environment involves one-on-one transactions to purchase goods or services, and often focuses on a relatively narrow view of reputation information only [6]. We have found that inter-enterprise collaborations, which involve multiple partners and a wide range of interdependent services, require a broader information model in order to capture the variety of risks and incentives as well as their dependence on the decision context [3].

Research on human information needs for trust decisions has been made both in general and in the aforementioned electronic commerce setting; we have surveyed the work and translated it into inter-enterprise collaborations [5]. As a well-known example, McKnight and Chervany have proposed a model of trust decisions that takes into account a variety of factors, ranging from a situational willingness to trust to a mental disposition, a belief that the target will behave in a trustworthy manner and even whether structural assurances make the situation appear less risky [7]. Different factors affecting trust decisions can be categorized based on whether they are specific to the trustor only (such as disposition), depend on the trustee (such as reputation), or the given decision context (such as the business importance of the action) [5].

We have gathered the aforementioned information requirements and evaluated the Pilarcos trust information model and trust decision process against them to ensure that the expert tool provides the information needed for human trust decision making [5]. Further input on the user interface design was gained from Nielsen's usability principles for designing user interfaces [8] and different cognitive strategies: Cognitive Fit Theory, Cognitive Load Theory, Unified Theory of Acceptance and Use of Technology and Technology Acceptance Model [9], [10], [11].

### III. SEMI-AUTOMATED TRUST DECISIONS IN PILARCOS

In this section, we first discuss trust management foundations in Pilarcos, and then present related work on trust-

aware inter-enterprise collaboration management.

#### A. The Pilarcos trust management system

The Pilarcos trust management system makes semi-automated, local and context-aware trust decisions. These decisions are repeated through the entire life cycle of the collaboration, whenever further resources need to be committed [1]. They compare a risk estimation, based on the past behaviour of the trustee represented as its reputation, against a chosen risk tolerance, which reflects the strategic importance of the tasks and the goals of the enterprise.

While routine trust decisions are automated to ensure the efficiency of collaboration management, there are always situations that require human intervention, for example due to a combination of high risk and high strategic importance. The local trust decision policies in Pilarcos therefore define risk tolerance ranges for automatic acceptance, automatic rejection, or for requesting a manual trust decision from a human user.

The Pilarcos trust management system makes automated trust decisions based on seven different parameters: trustor, trustee, action, risk, reputation, importance and context [3]. As discussed in the previous section, the trustor and trustee are business services operating within their respective enterprises. The action represents a collaboration task that needs to be performed, involving a commitment of the trustor's resources.

The risk and reputation factors are closely connected: risk estimates present probabilities of different outcomes, and reputation information is stored as experience counters of different observed outcomes so far. These experiences are gathered both directly through local monitoring, and as shared information through reputation networks [3]. While shared reputation information may be erroneous and is therefore locally evaluated for credibility, it provides a valuable extension to the first-hand information particularly in the case of actors that are not previously known, and actors who have recently changed their behaviour.

Instead of trying to capture all possible outcomes of all actions, Pilarcos represents their effects on assets. There are four high-level asset classes: monetary, reputation, satisfaction and control [3]. Probabilities of the outcome effects are presented on the scale of large negative effect, slight negative effect, no effect, slight positive effect and large positive effect that the action has been observed to have on that asset. This condensation of possible outcomes to a set of categories for affected assets improves the reusability and interoperability of reputation information across different enterprises.

The monetary asset denotes any resources that can be represented in monetary terms. The reputation *asset* reflects the trustor's own public relations, appearance in the media, and attitudes of their customers and partners towards them [3]; in contrast, the reputation *information* discussed

above concerns the past behaviour of the trustee. The need for security, privacy and other aspects related to autonomy are represented by the control asset. Lastly, the degree of fulfillment of the trustor's expectations by the trustee is represented by the satisfaction asset: it is used to measure whether the trustee tends to respect its agreements [3].

The reputation counters of observed outcomes are converted into a risk estimate by transforming the absolute numbers into ratios: essentially, 5 major positive experiences out of 10 total experiences translates into a probability of 50% of a major positive outcome for that asset. Relevant adjustments are made to accommodate, e.g., low-stake actions that cannot have a large monetary effect, and credibility-based weighting between local and shared reputation information.

The risk estimate is compared to risk tolerance formulae to determine the outcome of the decision. The risk tolerance formulae may be adjusted automatically according to the strategic importance specified for the action; this essentially represents the known outcome of a positive decision, such as not having to compensate other collaborators due to withdrawal from the collaboration during its operation. In the automated trust decision making process, the different factors are also subject to change by the context parameter, which manifests as conditional filters, or modifiers, of the data to allow temporary situational adjustments in the system. One example of a context modifier is insurance, which can apply to all actions in one specific collaboration and essentially eliminate the monetary risk involved.

#### B. Related Work

The TrustCoM framework [12] and the ICT infrastructure of ECOLEAD [13] are trust management systems for inter-enterprise collaborations. Their approaches differ somewhat from Pilarcos.

The TrustCoM framework [12], [14] performs trust decision making during the joining and continuation of the collaboration. In contrast to Pilarcos, TrustCoM makes trust decisions only when a new partner needs to be added or previous partner needs to be replaced, not routinely at each resource commit. The trust decisions are made based on reputation information measuring trustee capabilities, integrity and benevolence, in addition to functional definitions of the role, requirements of quality of service, cost and security. The TrustCoM framework also involves a user interface in the form of an eLearning portal in a scenario demonstrator [14], helping users find the best service suitable for them. In the general case, the design of trustworthy and secure user interfaces falls outside the scope of TrustCoM, although its importance is acknowledged.

In ECOLEAD [13], [15], trust decisions are made at two points: base trust is established during the entry into the ecosystem and specific trust evaluated when each inter-enterprise collaboration is set up. For base trust, all enterprises entering the ecosystem answer a questionnaire

on, e.g., organizational competences, prior successful collaborations, prior engagement in opportunistic behaviour, and adherence to technology standards and delivery dates. Collaboration-specific trust is established in a hierarchical manner, starting from the specification of objectives in terms of measurable elements. The ICT infrastructure of ECOLEAD also provides support for portlets for interaction with the users [15]. The trust prototype has a web and mobile portlet, providing a list of potential partners for collaboration, where the users can select those found most suitable for the task. Like TrustCoM, ECOLEAD does not focus on user interfaces for trust-decision making specifically.

Handling human intervention for semiautomated trust decisions remains an open research question in related work. Factors to consider include the phenomenon of human trust decision making, information requirements of the human users, the appropriate way of presenting the information, and reducing the frequency of calls for human intervention in the future, to ensure that the efficiency of collaboration management is maintained. We have summarized our earlier work [5] on the two former in earlier sections, and continue on the two latter in the following section.

#### IV. USER INTERFACE OF THE TRUST DECISION EXPERT TOOL

In this section, we present the design of the user interface handling human interventions on trust decisions. The presented user interface extends the Pilarcos trust management system.

In accordance with the Pilarcos trust information model, the user interface presents information about risk, reputation, goals affecting the importance of the action, and context [16]. Within its main information views, it presents further details on the credibility of the information, behavioural changes that can affect the validity of the reputation information, assets endangered according to the risk tolerance comparison, and the progress status of the collaboration when trust decisions need to be made during the operational phase of the collaboration. The information is presented as a combination of textual and graphical formats. Figure 1 shows an example risk view of the user interface; the other major views of reputation, context and progress information are minimized in the screenshot.

On the top, the user interface presents the goals of the inter-enterprise collaboration, such as earning money, gaining experience or building reputation. The importance of the goals that the enterprise has set for the collaboration encourage a positive trust decision. In addition to the goals, the deadline for making the trust decision is prominently shown. Both these information elements and their placement promote transparency.

The risk view presented in the figure shows the produced risk estimate, represented in the form of probabilities of



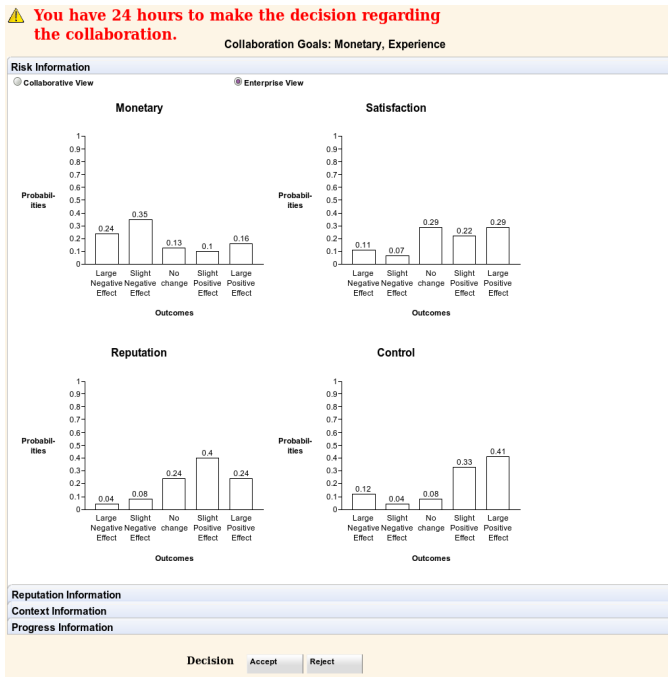


Figure 1. Risk information view of the trust decision expert tool.

different outcomes. These outcomes for different assets follow the trust information model of Pilarcos, as described in the previous section. The four asset classes correspond to four graphs in the risk information view.

The risk information can be studied through two different views: collaborative and enterprise view. The collaborative view presents collective risk probabilities for the collaboration as a whole; it reflects the fact that even though a trust decision is generally made concerning a one-on-one interaction within a larger collaboration, other participants in the collaboration may have a strong influence on the eventual outcome of the action. A manager may consider placing an order to a generally reliable contractor, for example, yet decide against it because it cannot trust its proposed subcontractors for this collaboration. In contrast, the enterprise view provides information about the risk posed by the single trustee individually. The current version of the trust decision expert tool presents the collaborative and enterprise view in the same format.

The reputation view provides background information to the risk estimate. Reputation information is presented as graphs. It consists of experiences, reflecting the past and present behaviour of the trustee on the same outcome scale as the risk information. The view also shows the estimated credibility of the shared reputation information, and presents whether the trustee’s behaviour has been consistent or not, which may have an impact on the validity of the available reputation information. Behavioral consistency is expressed through the number of times the system has detected a

change in the actor’s behaviour [4], and by showing both the overall experiences and the experiences based on the current period of consistent behaviour. Finally, the view uses colours to indicate the assets for which the risk estimate is not within the automatically acceptable risk tolerance bounds, as the actor’s reputation information for these specific assets may be of particular interest.

In automated trust decisions in Pilarcos, risk estimates are compared against risk tolerance, which is in turn based on the strategic importance of the action at hand. While the importance is represented through the goals of the collaboration, and tolerance constraints are partially visible through the assets shown not to be within limits, in manual trust decisions the human user is responsible for analyzing and setting the actual risk tolerance limits for the decision.

In the expert tool, the context information view presents the currently active context items to the human users through simple textual phrases, such as “Enterprise A is an important strategic partner” or “The current collaboration is covered by insurance”. This information is collected from the descriptive metadata of any active context filters.

Finally, the progress information view of the expert tool supports trust decisions on ongoing collaborations. The view presents the progress of the collaboration in graphical format, visualizing the tasks completed by different partners.

The eventual user decision is either to accept and approve the action or reject it, which generally results in a withdrawal from the collaboration. In addition, the tool will ask the user to provide a scope for the trust decision: whether it applies for the remainder of the contract, or for a given time period, or for this specific decision only. This helps reduce the frequency of requests for human intervention, as further decisions needed within the set scope can be automated. The scope information is stored as a context filter, which overrides the risk tolerance formulae appropriately to automatic rejection or acceptance for future decisions within the given scope.

## V. USER EVALUATIONS

This section presents the user study setup and initial results of the user evaluations of the trust decision expert tool [16]. We evaluate the interface from four points of view: (i) sufficiency of the presented information, (ii) usability, (iii) user performance and (iv) quality. All the participants are researchers more or less familiar with the Pilarcos trust management system. The main objective behind recruiting such participants is to gather feedback from users who are representative of the actual target user base of Pilarcos. The user study takes around one hour per user.

The user studies are conducted in three phases: introduction, solving test tasks and debriefing. During the introduction phase, test participants are introduced to the user interface and test setup. After the introduction by

Statement	St. Ag.	Agree	Neutral	Disagree	St. Dis.
Sufficiency of information	-	3	-	2	-
Ease of finding information	-	3	2	-	-
Clarity of presentation	-	4	-	1	-
Correlation btw. information presentation and tasks	2	1	2	-	-
Ease of use	-	3	2	-	-
Confidence of using	-	3	2	-	-
Willingness to use in future	-	4	-	1	-
Feel safe to use	-	-	4	1	-

Table 1  
SUMMARY OF THE USERS' OVERALL EVALUATIONS.

the moderator, the test participants are presented with the following test scenario:

*You are running an enterprise named “Quick Service”, which provides online logistic services within Europe. Your enterprise is involved in collaborating online with other enterprises throughout the world. You are using the Pilarcos middleware for managing your online collaborations. Usually, Pilarcos middleware makes automated decisions regarding your enterprise’s participation in the online collaborations, but now you have received an email, containing a link, asking you to make a decision regarding your continuation in an ongoing collaboration.*

Based on the test scenario, the participants are asked to write their expectations about information that they would like to have for making trust decisions in such a situation. After the introductory phase, the test participants are allowed to study the user interface themselves for getting familiar with it. Afterwards, the test participants are asked to perform the tasks using the user interface. The completion of each task is followed by a short questionnaire capturing the real-time experience of the participants after each task. The test participants are encouraged to think aloud while performing the test tasks. The “think aloud” methodology is employed for getting insights into the problems and thought process of the participants while they are performing the test tasks [8]. The moderator notes the participant comments while performing the test task as well as time taken by them to perform each task. Finally, during the debriefing phase, the test participants are asked to fill in a post questionnaire aimed at gathering general experience and impressions about the user interface. The user evaluations have been made with five test participants, which has provided us with quite useful feedback.

The first point of view evaluated the *sufficiency of the information presented* to the human users for trust decision making. As previously mentioned, the user interface presents risk, reputation, context, collaboration progress status, goals and credibility information for trust decision making. Table 1 shows the user rating of the user interface in terms of the sufficiency of the presented information. Based on the analysis of the debriefing phase and participant comments

while performing the tasks, we believe the probable reason for disagreement might be the absence of some relevant information, such as the value of the contract in terms of possible monetary profits to the enterprise. Another suggestion for enhancing the available information concerned a more detailed representation of the collaboration progress, relating it to the underlying business process model instead of simple milestones.

The second point of view of *usability* evaluated the ease of using the user interface in terms of ease of finding the presented information, clarity and existence of correlation between the information presentation formats and tasks to be performed. Missing or unclear information were again a probable cause for critique here. For example, the test participants were unclear about the ontological meaning of the assets. Furthermore, some users suggested that a summarized and concise view of the already presented information should be added, including, for example, small textual sentences such as “you have 63% probability of gaining monetary benefits”.

The third point of view of the *user performance* evaluates the user interface in terms of the success rate of task completion, number of errors committed while performing the tasks, and time taken for task performance. The evaluation results reveal that the task completion rate is 100% irrespective of accuracy. However, when considering the factor of accuracy, the successful task completion rate is 100% for only two of the participants. It is 93% for two other users, and 78% for one participant. In other words, the error rate is 7%. We suspect the lack of attentive focus while reading the tasks to be the main reason for the existing error rate, because we found the same participants giving correct answers for other similar tasks. Regarding task completion timing, we found that three of the test participants are able to perform 71% of the tasks within seconds, whereas the remaining participants perform respectively 79% and 93% of the tasks in seconds. The time taken is agreeable considering the novelty of the tool, as none of the participants have ever used any kind of trust decision expert tool before.

The fourth point of view of *quality* aims to evaluate the user satisfaction of using the user interface in terms of ease of use, confidence, willingness to use and perception about security. The evaluations are summarized on the last four lines of Table 1. As mentioned previously, insufficient presented information seems to be a likely reason for disagreements or neutral opinions.

In general, we found that test participants found the information presentation formats to be easy to read and understand. They also stated they found the user interface to be intuitive.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we have presented a user interface for trust decisions on inter-enterprise collaborations. The motivation

for this work arises from the fact that while we wish to automate routine decisions on collaborations, there are situations where human intervention is necessary. We have evaluated the implemented expert tool with the help of users familiar with the underlying Pilarcos system in order to collect feedback on improving the tool. We found that the current version of the user interface of the trust decision expert tool is appreciated by the test participants. However, the user evaluation also revealed a number of shortcomings, which provide significant pointers for not only improving the existing version of the tool, but also possible future research directions.

The user interface for simple trust decisions is a first step in a larger project to enable direct user interaction with the entire Pilarcos collaboration management toolset [2], which has so far focused on automating the relevant support processes. As a direct extension of the basic trust decisions, a planned collaboration negotiation expert tool should allow the user to simulate the possible outcomes of coming to a negative decision: Are there choices of better partners? Can the terms of the contract be adjusted before entering the collaboration? If a partner can no longer be trusted, can it be replaced? Would a different trust decision policy perform better for this collaboration, requiring fewer human interventions?

The work on the expert tool also unveils a need to extend the underlying information model. We concluded early in the design process that trustee-specific risk estimates must be complemented by a collaboration-wide risk estimation, and will further explore the strength of different methods of producing these collective estimates. During the user evaluation, it became apparent that human users are influenced by example, and therefore the relevant decision history (of both automated and human decisions) has an important role and should be visualized to the user. Similarly, while the textual descriptions of context filters have no effect on automated processing, they play an important role in the user interface. We will continue this work in parallel to the expert tool development.

#### REFERENCES

- [1] S. Ruohomaa and L. Kutvonen, "Trust and distrust in adaptive inter-enterprise collaboration management," *Journal of Theoretical and Applied Electronic Commerce Research, Special Issue on Trust and Trust Management*, vol. 5, no. 2, pp. 118–136, Aug. 2010.
- [2] L. Kutvonen, T. Ruokolainen, S. Ruohomaa, and J. Metso, "Service-oriented middleware for managing inter-enterprise collaborations," in *Global Implications of Modern Enterprise Information Systems: Technologies and Applications*, ser. Advances in Enterprise Information Systems (AEIS). IGI Global, Dec. 2008, pp. 209–241.
- [3] S. Ruohomaa and L. Kutvonen, "Making multi-dimensional trust decisions on inter-enterprise collaborations," in *Proceedings of the Third International Conference on Availability, Security and Reliability (ARES 2008)*. Barcelona, Spain: IEEE, Mar. 2008, pp. 873–880.
- [4] S. Ruohomaa, A. Hankalahti, and L. Kutvonen, "Detecting and reacting to changes in reputation flows," in *Trust Management V*, ser. IFIP AICT, vol. 358, Copenhagen, Denmark, Jun. 2011, pp. 19–34.
- [5] P. Kaur and S. Ruohomaa, "Human intervention on trust decisions for inter-enterprise collaborations," in *Post-Proceedings of the 15th IEEE International EDOC PhD Symposium*, 2011, to appear.
- [6] Y. Yao, S. Ruohomaa, and F. Xu, "Addressing common vulnerabilities of reputation systems for electronic commerce," *Journal of Theoretical and Applied Electronic Commerce Research*, 2011, accepted for publication.
- [7] D. McKnight and N. Chervany, "The meaning of trust," University of Minnesota, Minneapolis, MN, Tech. Rep. 9604, 1996.
- [8] J. Nielsen, *Usability Engineering*. Boston: M.A. Academic Press, 1993.
- [9] I. Vessey and D. Galletta, "Cognitive fit: An empirical study of information acquisition," *Information Systems Research*, vol. 2, no. 1, pp. 63–84, 1991.
- [10] L. Oshlyansky, P. Cairns, and H. Thimbleby, "Validating the unified theory of acceptance and use of technology (UTAUT) tool cross-culturally," in *Proceedings of the 21st British HCI Group Annual Conference on People and Computers*, vol. 2, ser. BCS-HCI '07. Swinton, UK: British Computer Society, 2007, pp. 83–86.
- [11] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257 – 285, 1988.
- [12] M. Wilson, D. Chadwick, T. Dimitrakos, J. Doser, A. Arenas, P. Giambiagi *et al.*, "The TrustCoM Framework v0.5," in *6th IFIP Working Conference on Virtual Enterprises (PRO-VE 2005)*, 2005.
- [13] S. S. Msanjila, H. Afsarmanesh, J. Hodik, M. Rehk, and L. M. Camarinha-Matos, "ECOLEAD deliverable D21.4b: Creating and supporting trust culture in VBEs," EC Information Society, Tech. Rep., Mar. 2006.
- [14] R. Ratti, M. del Mar Rodrigo Castro, C. A. Ferrandiz, S. Mores, R. Rabelo, R. J. T. Junior, and P. Gibert, "TrustCoM project final report," European Commission, Tech. Rep., 2007.
- [15] R. Rabelo, S. Gusmeroli, C. Arana, and T. Nagellen, "The Ecoload ICT infrastructure for collaborative networked organizations," in *Network-Centric Collaboration and Supporting Frameworks*, ser. IFIP International Federation for Information Processing. Springer Boston, 2006, vol. 224, pp. 451–460.
- [16] P. Kaur, "Supporting users' trust decisions on inter-enterprise collaboration," Master's thesis, Aalto University; Tech. Rep., University of Helsinki, 2011.

## Human Operator Perspective to Autonomic Network Management

Marja Liinasuo, Iina Aaltonen, Hannu Karvonen  
 VTT Technical Research Centre of Finland  
 Espoo, Finland  
 e-mail: [marja.liinasuo@vtt.fi](mailto:marja.liinasuo@vtt.fi), [iina.aaltonen@vtt.fi](mailto:iina.aaltonen@vtt.fi),  
[hannu.karvonen@vtt.fi](mailto:hannu.karvonen@vtt.fi)

Beatriz Fuentes, Alfonso Castro  
 Network Oriented Automation Department  
 Telefónica I+D  
 Madrid, Spain  
 e-mail: [fuentes@tid.es](mailto:fuentes@tid.es), [acast@tid.es](mailto:acast@tid.es)

**Abstract** - Autonomic Communications is a promising approach for the management of future networks, based on providing self-management and other self-x capabilities to the network elements. Network Governance approach aims to develop new methods for operators to efficiently manage this intelligent infrastructure, ensuring human-to-network communication. This paper presents an interview-based analysis of the expertise of human network operators in their network management activity and the derived requirements for the governance of a self-managed network from a human point of view. The requirements cover the aspects of trust, division of labor between human operators and self-x functionalities, self-x design and human-to-network interface.

**Keywords** - autonomic networks; self-x; Network Governance; human operators; human factors.

### I. INTRODUCTION

Future computation, storage, and communication services will be highly dynamic and ubiquitous: an already increasing number of heterogeneous devices will be used from different places to access a myriad of very different services and/or applications. Users' devices, smart objects, machines, platforms, and the surrounding space will be interconnected as a decentralized tree of resources, conforming dynamic networks of networks. The complexity of managing such a network presents a challenge for telecom operators.

Current operation support paradigms cannot effectively manage this situation; rather a new reliable, dynamic, and secured communication infrastructure with highly distributed capabilities is needed. The autonomic network approach envisions meeting these features, helping network operators to achieve the desired levels of dynamicity, efficiency, and scalability to manage current and future networks.

Autonomic computing was first introduced by Paul Horn as a solution to the increasing complexity in the management of IT systems [1]. The suggestion was to design and build future systems and infrastructures capable of running themselves and adjusting to varying circumstances by taking the massively complex systems of the human body as a model. In fact, the name "autonomic" was derived from the human autonomous system. This proposal was well accepted by the industrial and scientific community, and it was extended to other fields. Then, Autonomic Communications was born with the objective of providing autonomic behavior to network infrastructures [2][3]. The terms self-monitoring,

self-diagnosis, self-configuration, self-healing, and in general self-x were introduced to denote the new capabilities of the network nodes to manage themselves without external intervention.

UniverSelf [4] is an EU-funded FP7 project with the aims of overcoming the growing management complexity of future networking systems and reducing the barriers that complexity and ossification pose to further growth. The project aims at consolidating autonomic methods of the future Internet into a novel Unified Management Framework. It also includes the design of a privileged, powerful and evolved human-to-network interface that will be used by the human operator for expressing their business goals and requests, thus shifting from network management to network governance. Network Governance approach is meant to provide a mechanism for the operator to adjust the features of the demanded service/infrastructure using a high level language [5]. These high level directives must be translated into low level policy rules that can be enforceable to control the behavior of the autonomous agents.

Since the new management or government framework is intended to be used by human network operators (HNOs), it is of utmost importance to produce solutions that are usable by human users and meaningful in the context of their work. Presently, it is not known how autonomic functionalities should be designed in order to be functional from the human operator's viewpoint. Thus, it is important to learn what the characteristics of the HNO work are from the perspective of these professionals and what the specific demands are that the work sets on human performance. Also the conceptions regarding autonomic tools should be elaborated. For this purpose, as part of the UniverSelf project, we have carried out a set of interviews to human operators of two European telecom companies. This paper describes the results of these interviews. In particular, HNOs' views of the network characteristics and network management, their knowledge on autonomic functionalities, and the impacts (both benefits and risks) they foresee after the potential deployment of a self-managed network are focused on.

The rest of this paper is structured as follows. Section II presents a deeper elaboration on the objectives of this work. Section III details the methods utilized in the interviews and the analysis of the answers. Section IV presents the results of this analysis, which are further refined in the discussion included in Section V. This section presents the requirements that were extracted from the interviews and grouped into

requirements on network maintenance and requirements on self-x functionalities. Finally, Section VI concludes the paper by summarizing the outcomes of this work and anticipating future work.

## II. OBJECTIVE

The tremendous evolution of network technologies in terms of their design, capabilities, and capacities has not been accompanied by advancements of the same magnitude in network management solutions [6]. Therefore, network management depends currently on the abilities of HNOs, and it is still a widely unknown area with very little knowledge from the human network operators' perspective. The willingness of HNOs to accept the new autonomic networks may constitute one bottleneck in the transitional period from human-centered to autonomically manageable networks. In this direction, when a new technological solution such as a self-managing network becomes ready to be deployed in large scale, the role of the HNOs, the demands of their work and other possibly affecting factors should be addressed in order to guarantee a smooth transition phase from deployment to flexible and efficient every-day use.

An analysis showed that the strategy to manage network operability depends on the demands and possibilities of the situation that vary in the continuums of options available for network operations and the instability of the situation [7]. In this study, however, the key personnel that had access to both strategic and operational level was found to be middle level professionals. Regarding network operation, it has been claimed that the cause of a major failure in system operation is not an accident but a conscious choice that proves wrong in that specific situation [8]. This indicates that when studying network operation in an operational level, the whole working system must be taken into account. In Finland, a study has been performed about the dependability of IP networks, including also human factors research related to the meaning of human error in the work of a HNO [9][10]. It was found out, among other things, that the error resulting from the work of HNO may originate from a lapse or typo but may also be a side effect of the whole working system that has deficiencies in its functioning.

Our approach in the present study was to seek information by enquiring into the work of technical experts that are currently operating the network in different companies. The main objective was to get information on the work of the HNO and additionally on user acceptance, i.e., the factors that effect and ease the deployment of autonomic network solutions from the HNO point of view. The aim was to get information directly from the professionals who operate the network, avoiding managerial levels. This practical approach provided information about the insight of the user perspective on network operation. In short, interviews were double-faced: on one hand, they tried to get a picture of the general network management as perceived by a HNO; on the other hand, there was a set of questions aiming to extract their conceptions about self-x functionalities.

The objective of this document is to present the results, analysis process, and conclusion of the HNOs' interviews.

From this analysis, general requirements on network automation can be extracted. These requirements can be used for guiding the design and implementation of the governance of future networks.

## III. METHODS

The background approach of the interviews was based on Core-Task Analysis method [11]. The aim of Core-Task Analysis is to identify the core task of a specific work. Core task is the main result-oriented content of the work that can be derived by analyzing the objective of work and the demands that the objective lays on workers both in general and in specific situations.

Interviews were conducted in telecommunication companies participating in the UniverSelf project. The interviews focused on the perspective of a HNO performing network monitoring or some other corresponding work in the same level, lower than the one of a manager, so that the characteristics of the network as a target of work would become highlighted.

Interviews were conducted anonymously in the native language of the interviewee.

### A. Data gathering

All but one of the interviews were performed in a meeting room at the company premises where only the interviewer and interviewee were present. One interview was performed over telephone.

The interview comprised of 40 questions. The questions were built on two main themes: a) work characteristics of a HNO as conceived by the human operator and b) opinions of the manageability of self-x functionalities. Interviews were conducted during April and May 2011. Results of 17 interviews, acquired from two European telecom companies, are presented. The duration of interviews ranged from 10 minutes to 1 hour 40 minutes, which is reflected in the variety of quality in the interview data.

All interviews were audio recorded. The recordings were later on transcribed and translated into English for analysis purposes, resulting in about 180 pages of interview material.

### B. Analysis

The performed analysis was qualitative. As the interviewees did not form a representative or statistically significant sample of all HNOs, all replies are important. Thus, even if an opinion expressed by several interviewees was regarded as more general and was self-evidently taken into account, single replies may have acquired similar weight in the analysis process if the expressed conception included a clearly expressed idea.

The analysis took place in several phases. First, the transcribed interviews were grouped according to interview questions. Similarly, questions that needed to be scrutinized together as well as questions that produced similar replies were grouped together. Then, the core idea or main message in each interview reply was extracted. These ideas or messages are here called notions. In practice, a notion is usually a shortened version of the expressed opinion (reply). In some cases, interviewee presented more than one notion

in the reply; then, the notions within the reply were categorized in more than one category. Thereafter, similarities among the notions were sought for. Investigating iteratively the notions and the possible similarities among them, similarities between notions became gradually more apparent and notion categories could be created. All the notions belonging to the same category have about the same core message from the perspective of that category.

When extracting the results from the analysis, several information sources, based on interview data, were simultaneously available and used:

- the classified notion categories and the notions themselves, collected question-wise
- the original replies for each question, collected question-wise
- the replies by each interviewee, collected interviewee-wise.

The notion categories were important in acquiring an overview of the replies. This information source was the most important due to the amount of information. Original replies were important if the validity of some specific reply had to be checked. The interviewee-wise information is important to know if some deviant or specific opinion is presented; for instance, if the interviewee had some specific occupation, (s)he may have produced replies different from other interviewees' replies due to this deviant working experience. When reporting the results, mainly notions and notion categories are used.

The interview data were classified according to general topics: (1) background information related to the work of the interviewees, (2) network characteristics as perceived by human operator, (3) interviewees' work with the network, and (4) conceptions of self-x functionalities.

Background information (above, class 1) was used for validating the results so that it is known what is the source of the opinions and information gathered in the interviews.

Network characteristics and interviewees' work with network (above, classes 2 and 3) already tell what the work demands and difficulties are in network operation. Thus, even if the interviewees may not be expressive in the answers related to self-x functionalities, the demands the network sets for the human operator were expressed here. Work demands and difficulties revealed in these replies can be used for inferring in what way self-x could support network maintenance.

Opinions of self-x functionalities (above, class 4) tell directly how human operators assume the self-x could support network maintenance and what should be avoided when designing them.

Finally, the conceptions classified as above were interpreted in the light of the requirements the domain sets to the work and tools of HNO from the two perspectives: a) requirements on network maintenance, and b) requirements on qualities of the self-x functionalities.

#### IV. RESULTS

This section summarizes the result of the analysis of the interviews. Requirements were extracted on two main

topics: requirements on network maintenance and on self-x functionalities.

##### A. Requirements on network maintenance

The requirements the network maintenance sets for HNOs are based on the uncertainty and complexity of network for many reasons: (i) technically (30 notions), a network is affected by many factors that are hard to control (especially complexity due to various types of equipment and the variety of manufacturers). Obsolete, but functional systems create a challenge in themselves as their use prevents the transformation of the network to become more homogeneous and easier to maintain. Furthermore, (ii) a network is never perceived directly but through supervision mechanisms that also may fail or provide insufficient information that cannot be trusted (5 notions). Also, (iii) human work is not always perfect (27 notions): information is not always delivered in time, it may not have been registered in the system that is used in network maintenance, and there might be hurry due to customer needs. Finally, (iv) weather (5 notions) and physical problems (accidental cable cuts and the stealing of cables; 6 notions) cause trouble that are hard to control.

The criterion of the good status of network is clear and strict. Either directly or indirectly (via alarms sent by the supervisory system) the availability and functioning of the network as perceived by the end user or customer service was the most usual way to evaluate the seriousness of a network problem (15 notions, opposed to 6 notions where the operators only referred to their own work). Depending on the Service Level Agreement (SLA), problem must be solved within a certain period.

Network maintenance requires mastering a large set of knowledge related to the various equipment and manufacturers and the ability to react rapidly to network breakdowns, knowing that each breakdown could be solved in a different manner from the previous one. HNOs cannot rely on the information acquired earlier but must update their knowledge constantly in order to be able to maintain the network functional. The demands of acquiring enough information and of maintaining it during rapidly changing situations are hard. The piece of information acquired earlier may be unavailable as the next time it is needed is several months later.

Also cooperation and communication are needed. Network is in a constant change and information of these changes should be constantly updated so that everybody has the essential information available. Additionally, cooperation is an essential part of work for most HNOs; for some, cooperation is required to get help and support when needed (6 notions) and for others, cooperation is part of normal work procedures (11 notions).

##### B. Requirements on self-x functionalities

The design of the self-x functionality from the perspective of a HNO depends on what the role of this functionality has relative to the tasks that the human operators have on their responsibility. The interviewees, HNOs, did not have a clear vision what the options in this

kind of situation could be, nor were they asked to further elaborate a vision of such a possible future. In many cases, HNOs seemed to describe present functionalities that are automatic instead of elaborating the features of future autonomic functionalities. Some clues can be found from the replies. First, interviewees were aware that self-x functionalities could reduce the amount of personnel needed (22 notions). Second, some opinions were presented according to which the work of the human operator could become easier or that the deployment of self-x results in the possibility for the HNO to do some other, but more productive work (18 notions). Third, self-x could perform better (quicker etc.) than presently (12 notions).

In the following, the general requirements are described, drawn on the information gathered from the interviews.

#### 1) *Trust*

The trust of the human operator to self-x can be easily destroyed by erroneous functioning of the self-x functionality in question. Possibly, any or only a very few mistakes is sufficient to ruin trust and willingness to use the self-x functionality (16 notions). Furthermore, if the functionality is of the kind that diminishes the need of human personnel, it is probably more intolerable for the rest of personnel to work with such functionality.

The interviewed HNOs stated that demanding tasks – which require human ingenuity as well as solving new and unexpected problems – cannot be performed autonomically (9 notions). The self-x functionalities in this context should work without problems that require human intervention: such problems would probably pose the greatest challenge in network maintenance from the human operator point of view.

Based on the perceived benefits of self-x expressed in the interviews, especially performance speed (7 notions), smaller amount of network breakdowns (3 notions), and stability of performance (10 notions) are characteristics that are valued among HNOs. In addition, the possibility to dedicate routine tasks to self-x functionalities was found an advantage (15 notions). If these qualities can be provided and they function practically without faults, the self-x could be perceived as valuable and needed by HNOs maintaining the network.

#### 2) *The need of human control*

In the interviews, it was repeatedly stated that the human operator must have some control over the self-x functionalities: the lack of control was perceived as the greatest danger in autonomic functionalities (8 notions). Fault and failure of the autonomic functionality are important to know especially if the HNO needs to act in order to solve the problem. HNOs must also have enough information of the autonomic process so that they can intervene manually in a meaningful way, for instance to perform the task at least partly in lieu of the self-x if needed. If the failure of the self-x causes trouble for network maintenance, amending measures should be easily performable. In general, human control is needed for knowing that the functionalities are working; the urgency to get this status information and the need to control or maintain further the self-x functionalities depends on the nature of the task in question.

#### 3) *Interface between HNO and self-x*

The performance of self-x functionalities should be visible to the human operator when the self-x performance is somehow connected to the tasks of the human operator. HNOs need information of the functioning of the self-x in a general level. The work of the human operator requires sometimes rapid reactions; it is preferable to deliver clearly and briefly only the most important information, and, when needed, the human operator should be able to find more information about the matter relevant in the context in question.

The interface should include information that enables the evaluation of the performance of the self-x functionality so that it becomes evident whether it is functional or not. If possible, the system should deliver information that supports the HNOs in fixing the problem. In some cases it might even be necessary for the HNO to do manually the work or part of the work that was in the responsibility of the self-x. Last but not least, autonomic system should not overload the human user with recurring information. All in all, opinions of the good interface qualities were scarce and scattered.

#### 4) *Designing self-x functionalities*

The domain of telecommunication seems to be complex and hard to control. On one hand, operators are somewhat used to the fact that the tools they are using are not perfect at least in the beginning of their deployment. On the other hand, operators should not be burdened by poorly functioning tools as the work is quite straining already and because weaknesses in network maintenance are immediately reflected in the quality of services and customer satisfaction, vital for the telecom companies. Before taking a new functionality into use, it should be thoroughly verified that the functionality works; redesign might be needed later if changes affecting this specific self-x are made in the network.

#### 5) *Human tasks*

Tasks that are obviously left for humans are all physical tasks: implementing new cables and equipment, organizing cooperation among humans when a cable is cut or equipment must be replaced, and each time something unexpected or totally new is introduced to the network. Some human control must also be maintained to supervise the functioning of the network and of the self-x. What are the tasks that can be totally or mainly relied on self-x functionalities and when and to what extent self-x functionalities may require human involvement are questions that perhaps should be solved by each telecom company, depending on the kind of supervising tools and network equipment they have. Furthermore, the quality of the task determines the party performing the task; for instance, tasks requiring decision making in a highly specific situation without predefined preferences might best be left for humans whereas those requiring rapid and complex functioning that can be performed with existing definitions can be left for autonomic functionalities.

## V. DISCUSSION

This paper has presented the methods, analysis, and results of the interviews of HNOs carried out in UniverSelf

project. The interviews have shown that network operators define the network as complex, uncertain, and hard to control. Therefore, specific knowledge is needed to handle the issues that may appear. An autonomic network aiming to be deployed should embed this knowledge in order to success in its self-managed activity.

The analysis has also highlighted that human operators see self-x functionalities as supporting their current work; not necessarily replacing them or performing independently. Even when some of the interviewees expressed their fears about personnel reduction in case autonomic functionalities are deployed into the network, the majority expressed a positive opinion about the introduction of self-x features. In this direction, the HNOs thought that their role might change from current active monitoring and problem solving to an expert performing higher level tasks and solving only highly complex or unexpected problems, while routine tasks are left to the autonomic network. On the other hand, it also became evident that there was no deep understanding of the nature of self-x functionalities among HNOs: The conception of autonomic functionalities was often mixed with the automatic functionalities that are presently used in telecom companies.

The conceptions of human operators on the impact of self-x functionalities – loss of work, division of difficult tasks to autonomic functionalities, and the emerging possibility to do more productive work themselves as autonomic functionalities take over routine tasks – can be translated into a vision in which there are three options for the shared work between HNO and self-x (see Figure 1).

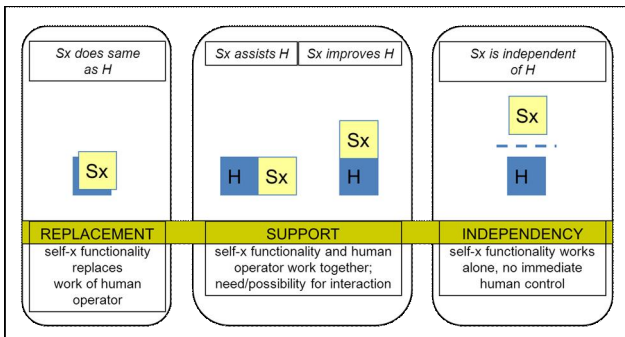


Figure 1. A graph describing the role of self-x functionalities relative to the role of the HNO. Yellow box “Sx” refers to tasks performed by self-x functionalities and blue box “H” refers to tasks performed by HNO. The figure describes three basic alternatives for a share of work between human and technology: Replacement, support and independency.

The conception of losing jobs with the deployment of self-x can be translated into such a solution where self-x functionality replaces human work in some specific task (in Figure 1, the leftmost option, “Replacement”). The option of the self-x functionality to facilitate the work of the HNO is not as easily translatable to a division of work between human and technology. Apparently, self-x functionality then assumes some subtask that the human used to perform or extends it further in some way (e.g., by performing more complex calculations than the human operator is able to do

within such a short period of time), supporting the work of the human operator. The functionality may perform part of the work human did earlier or the functionality could do something related to the task human performs but going beyond it for instance in speed or complexity (in Figure 1, the option in the center, “Support”, and there the left and right sub options “Sx assists H” and “Sx improves H”, respectively).

Interviewees thought they could devote themselves to more productive work while the self-x perform something else, possibly the tedious tasks that require a lot of mechanical work of the human operator but not so much human intellect. Obviously, here it is a question of work that the self-x functionality does entirely or mainly by itself. Basically, this situation develops whenever a self-x functionality takes over any part of human work as it then leaves the human a possibility to do something else, applying both “Replacement” and “Support” options.

Additionally, self-x could perform something that human is not able to do but that could either ease the HNO’s workload indirectly by, e.g., diminishing the amount of network failures, or elevating the functioning of the network in a totally new way. For this case we invented a third type of solution for the division of the work between the human operator and self-x functionality, labeled here as “Independency” (in Figure 1, the rightmost option). It illustrates the situation where the tasks of human and technology are separate. This option is valid in a situation where the self-x performs something that has never been on the responsibility of a human operator: it could be a new task that enhances or supports network availability or the like, in a new and unique manner.

Human control over the self-x functionalities is needed. If the self-x functionality is used for replacing human work (case “Replacement” in Figure 1) or it assists or improves human performance (case “Support” in Figure 1), it is important to know about the status of the autonomic functionality to guarantee a smooth joint performance between the human and autonomic functioning. More specifically, HNO needs to know if the self-x is functional; due to the nature of work, this should be evident to HNO. HNO also needs to know how to fix a problem in self-x functioning in order to keep the network manageable and functional (provided that the malfunctioning of the self-x functionality causes trouble in these areas). Furthermore, HNO should be able to know how to perform the same or about the same as the self-x functionality in case of failure of the functionality.

However, if the self-x is independent of human work (case “Independency” in Figure 1), the need for human control is not straightforward. An independent task is of the kind that is beyond human capabilities – hence, it can be concluded that when it works as defined, it raises the quality of network functioning but if malfunctioning it might cause trouble and is probably hard to control. Hence, HNO or at least an expert should be able to evaluate whether such functionality is functional but only an expert might have sufficient knowhow to repair a problem in such self-x



functionality. There is no possibility for a human to perform the same task manually, anyway.

Making amending measures requires, of course, that self-x functionality has an interface towards the HNO and that the human operator knows enough of the functioning of the self-x so that the information it delivers can be interpreted correctly. Thus, if the malfunctioning of the self-x causes trouble in normal network functioning and/or complicates the work of the human operator, HNO should be able to perceive that there is a problem in the self-x functionality and to infer the severity of it in order to decide whether to intervene or not and if to intervene, how urgently. Furthermore, it should be made possible to conclude what is/are the available way(s) to proceed in problem solving and to have access to means to perform the amending measures.

All the options presented above – replacement, support and independency – don't automatically result in a considerable decrease of personnel. It is possible that in addition to, or instead of decreasing the amount of personnel, self-x functionalities could enable more efficient network maintenance and higher-level services. HNOs could operate alongside self-x or only by human power alone while the more elementary issues are left for the responsibility of self-x. To sum up, the benefits self-x functionalities could bring are not only cost reduction in the form of decreased amount of personnel, but also higher income to the telecom operator due to more sophisticated, more stable, or novel services to customers. What is the relative impact of each of these possibilities remains to be seen.

## VI. CONCLUSION AND FUTURE WORK

How soon these visions of autonomic networks become true depends on the rate of emergence of self-x functionalities in network maintenance on large scale and on the qualities of the self-x. Possibly some type of joint work arrangement between the human operator and the self-x dominates at first, and the situation evolves with time.

To be effectively useful and widely adopted, the new management system, which also includes the management of self-x functionalities, must provide the telecom operators with tools to seamlessly govern the network infrastructure by means of decision oriented operational tasks rather than low level command execution. At the same time, reliable mechanisms should be included for the human operator to receive all the needed information for the supervision of the network, so that any impossibility to continue self-managed operation or realistic danger will be reported to humans with pertinent details of the situation. Furthermore, tools to start the recovering process are needed as well. The adoption of this approach should decrease the human intervention required for deploying new services and configuring and operating the network.

The results of our interviews have provided a number of requirements about network maintenance and self-x functionalities. Since almost all the knowledge required to manage the current networks resides in the human operators, they may constitute a bottleneck when trying to deploy an autonomic network. The generation of self-x functionalities

fulfilling the operator's demands and suggestions will become an important factor for the success of the autonomic networks. The research presented in this document supports the identification and extraction of operator's requirements for this purpose.

## ACKNOWLEDGMENT

The research leading to these results has been performed within the UniverSelf project ([www.UniverSelf-project.eu](http://www.UniverSelf-project.eu)) and received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 257513. Additionally, we thank the two telecom companies for giving the permission to perform the interviews and for making the practical arrangements for it; Helena Rivas and Martin Varela for kindly performing the interviews; and last but not least, all the human network operators who shared their opinions and conceptions with us in the interviews.

## REFERENCES

- [1] Horn, P., "Autonomic computing: IBM's Perspective on the State of Information Technology," 2001. Available at [http://www.research.ibm.com/autonomic/manifesto/autonomic\\_computing.pdf](http://www.research.ibm.com/autonomic/manifesto/autonomic_computing.pdf), accessed 2011.11.25.
- [2] Smirnov, M., "Autonomic Communication, Research Agenda for a New Communication Paradigm," Autonomic Communication, White Paper, 2004. Available at [http://www.autonomic-communication.org/publications/doc/WP\\_v02.pdf](http://www.autonomic-communication.org/publications/doc/WP_v02.pdf), accessed 2011.11.25.
- [3] Schmid S., Sifalakis M., and Hutchison D., "Towards autonomic Networks," Proc. of 3rd Annual Conference on Autonomic Networking, Autonomic Communication Workshop (IFIP AN/WAC), Paris, France, pp. 25-29, 2006.
- [4] <http://www.univerself-project.eu/>, UniverSelf webpage, accessed 2011.08.13
- [5] Chaparadza, R., Vigeriaux, M., Lozano-López, J. A., and González Muñoz, J. M., "Autonomic Mobility and Resource Management Over an Integrated Wireless Environment – A GANA Oriented Architecture," Proc. of the 2nd International Workshop on Management of Emerging Networks and Services (MENS), pp. 565-570, 2010.
- [6] Samaan, N. and Karmouch, A., "Towards Autonomic Network Management: an Analysis of Current and Future Research Directions," IEEE Communications Surveys & Tutorials, 11(3), pp. 22-36, 2009.
- [7] Schulman, P., Roe, E., Eeten, M., and Bruijne, M., "High reliability and the Management of Critical Infrastructures," Journal of Contingencies and Crisis Management, 12(1), pp. 14-28, 2004.
- [8] Besco, R. O., "Human performance breakdowns are rarely accidents: they are usually very poor choices with disastrous results," Journal of Hazardous Materials, Vol. 115, pp. 155-161, 2004.
- [9] <http://iplu.vtt.fi/>, IPLU web page, accessed 2011.08.26.
- [10] Norros, I. and Norros, L. "Human factors in the dependability of IP networks," ENISA Quarterly Review, 5(3), pp. 20-21, 2009.
- [11] Norros, L., "Acting Under Uncertainty – The Core-Task Analysis in Ecological Study of Work," VTT Publications, Espoo, 2004, pp. 241.



# Contents Enforme: Automatic Deformation of Content for Multi-features without Information Loss

Hiroaki Tobita

Sony Computer Science Laboratory Paris  
Paris, France  
tobby@csl.sony.fr

**Abstract**—We introduce a deformation technique that enlarges the feature areas in an image while retaining the information in the non-feature areas. Our main purpose is to provide an effective thumbnail that is useful for practical use in small display devices (e.g., cellular phones, digital cameras, and game devices). Even though many approaches to achieve our purpose have been developed, they are not useful because they require enough time to calculate features. In contrast, our approach can quickly deform image features based on a rapid segmentation technique that our laboratory has proposed. Moreover, our system calculates each segmented area, so it can treat multi-features as deformation elements. As a result, the feature areas were enlarged and the non-feature areas were reduced at almost processing speed, and the total information contained in the original image was retained. Our smooth deformation technique is useful not only for image deformation, but also for a wide variety of contents such as net-meeting and video contents. In this paper, we describe the concept underlying the image enforme technique and its applications.

**Keywords**-Contents Deformation; Information Retrieval; Video Compression; Zooming; Thumbnail; Net-meeting.

## I. INTRODUCTION

The continuing improvements in computer hardware have led to improved and more compact devices (e.g., cellular phones, personal digital assistants, and digital cameras), thereby making it easier for people to record and carry a huge amount of content, such as images, video files, and audio files. These devices are characterized by small displays with high resolution, so users can easily obtain information and examine the contents in detail. In addition, many content formats have been developed with a focus on effective file size. However, even though there have been considerable advances in device styles and media formats, the methods for previewing contents have remained almost unchanged. For example, to browse a set of images, a user has to look at a set of simple thumbnails or to look at the images individually.

To solve this viewing problem, considerable effort has been made towards providing effective thumbnails. Cropping [1, 2] is an effective way to create a clear thumbnail. The non-feature areas are cropped, leaving only areas with features. The resulting picture is clearer, and the thumbnails produced are useful for browsing images in small devices. Since the calculation involves mainly identifying features, it is very simple. However, cropping the non-feature areas eliminates some of the image's initial information, so the total amount of information is reduced. This makes it

difficult to distinguish between similar images. For example, if the original images contain similar objects but different backgrounds, the cropped images will be almost the same, making it difficult to retrieve a desired image.

Image retargeting is also an effective approach to emphasize an image's feature. Setlur et al. [3] developed an image deformation technique that uses a non-photorealistic algorithm. The objective is to provide effective small images by preserving the ability to reconcile important image features, so the quality of the features is an important element. Foreground objects are first separated from the background, and then a new smaller background image is created. After the background area is reduced, the foreground objects are restored to the image. As a result, the blank areas of the image are removed, and an effective small image that has the same-sized features is created. However, this technique focuses on a whole clear object, so the deformation scale is quite limited. Feng et al. [4] also provided a deformation technique that is based on a fisheye view [5]. In a deformed image, a feature area can be enlarged by liner scaling and non-feature areas can be reduced by non-linear scaling. However, this technique can only treat one feature area, so other features are not deformed and will be reduced if an image has multi-features. Most images contain multi-features (e.g., "two people" or "a person and a building"), so this technique has practical limitations. Moreover, these two approaches require enough time to calculate features, so they are impractical for small display devices that do not have enough calculation power.

We have developed a technique that supports effective content viewing [16]. We think that the following two elements are important for image deformation. First, consideration should be given to the relationship between the feature and non-feature areas, as Feng proposed [4]. The feature area is important for finding an image because it becomes the trigger for browsing. The non-feature area is also important because it becomes an element for recognizing similar images. Second, multi-features should be supported, as Setlur proposed [3]. An image generally has more than one feature, so the technique should deform such features clearly. Based on our approach, the deformed image contains advantages outlined by Setlur [3] and Feng [4].

We also think that calculation speed is important for small devices. Our main purpose is to provide an effective thumbnail for image browsing in small devices, so efficient calculation to deform image features is important because the calculation power of a small device is quite limited. The calculation process is divided into finding regions that users

focus on and deforming these regions. As we have also proposed a rapid image segmentation method [7], we used this method to retrieve features rapidly. The deformation depends on the result of image segmentation, so we treated each part of an object (e.g., only a face area or part of a building) as a feature area. To deform multi-features, we used rubber sheet [10] and constrained fisheye [11] techniques, thus, the deformed image had clear multi-features.

As a result, the deformed image had bigger feature areas and smaller non-feature areas than normal images. Although the gradation of the image changed, the total amount of information contained in the image remained the same. These processes were rapidly and automatically performed, so complex user interaction was unnecessary. A unique thumbnail was provided for use with small devices on the basis of setting deformed images.

The next section describes the image enforme overview. Section 3 describes the implementation of the system in detail. In Section 4, we mention the applications of our concept. In Section 5, we discuss advantages and limitations of our system. Section 6 concludes the paper.

## II. IMAGE ENFORME OVERVIEW

Figure 1 shows example deformed images, of which the retargeted image (1) had the highest quality. With image retargeting, however, if the foreground object itself is big, the deformed features are of the same size as the input features. Thus, the deformed image is not effective for information retrieval. The fisheye warp image (2) had a clear contrast between feature and non-feature regions. With fisheye warping, however, if an image has multi-features, the system can only recognize one region as a deformation area. Thus, second or third features are no longer useful.

An image produced using our deformation technique is shown in Fig. 1 (3). Using our approach, even part of an object can be a feature area, for example, if the image shows the face and body of a person and only the face area was deformed. Thus, in this example image users could identify the man's face more clearly than they could with the image shown in (1). Also, our approach treated multi-features as deformation areas, so the multi-features were more deformed and clearer in (3) than in (2). As we described, the non-feature areas were reduced and retained in the deformed image. Thus, users could easily understand what the subject was wearing as compared to the cropping method. Our objective is not to obtain excellent image quality but to produce clear features for effective thumbnails.

## III. IMPLEMENTATION

As shown in Fig. 2, there are two processes in the deformation step: extracting the feature areas from an image and deforming the feature areas. An image generally contains characteristic objects (faces, buildings, symbols, etc.); so the system retrieves these kinds of objects and then deforms them.

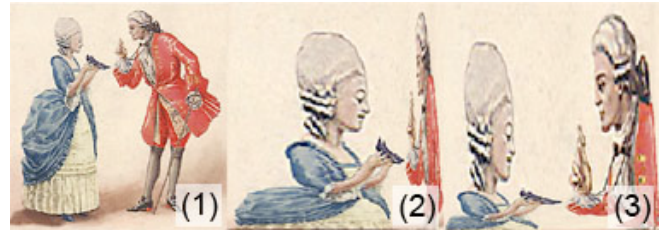


Figure 1. Example images deformed by image retargeting (1), fisheye warping (2), and image enforme (3).

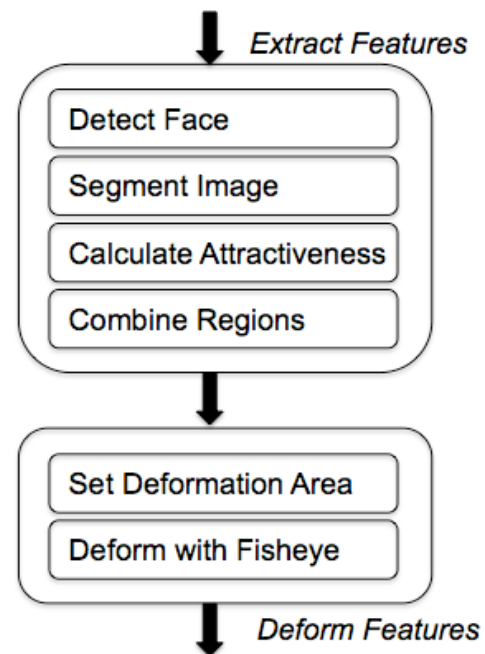


Figure 2. Deformation step is grouped into two processes: extracting feature areas and deforming them.

### A. Face Detection

We think that a human face in an image is the most important element, because the image itself would become strange if the system did not recognize a face as a feature, as shown in Figure 1 (2). Thus, a face detection process is first performed on an image. We use the method described by Sabe and Hidai [6] to detect the faces.

### B. Image Segmentation

After face detection, the system starts an image segmentation process. For this we use our previously proposed image segmentation method [7], which is a practical and readily available method for segmenting images (Fig. 3 (top)). It segments images into various regions; examples of the segmentation are shown in Fig. 3. After image segmentation, the system deletes the background regions using the method proposed by Tanaka [9]. In this method, the system first calculates the border between a focused region and its neighboring regions, then determines the foreground using the relationship between the length of the border and the length of the contour of the focused region.

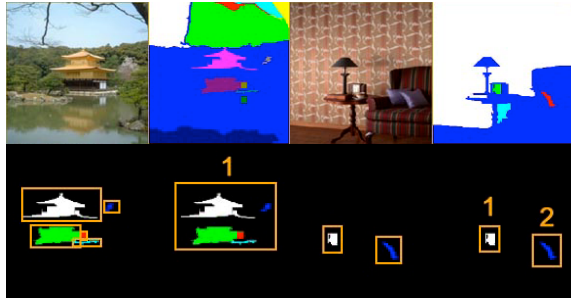


Figure 3. Extracting features: System retrieves feature areas through image segmentation (top). If the distance between regions is less than the threshold, regions are merged as a new region (bottom).

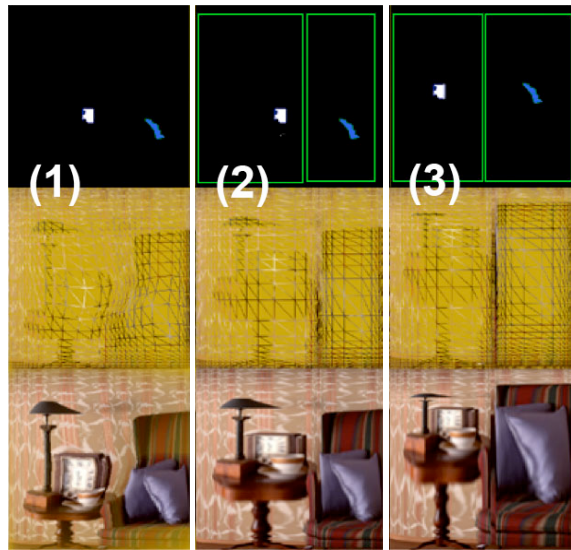


Figure 4. Comparisons of deformations: Images are deformed by a normal fisheye (1) and our multi-fisheye method (2, 3).

C. Attractiveness

To identify features, we need to know which region is important. We used Kansei factors [8] to identify highly attractive regions. The Kansei factor for each area was based on simple parameters (e.g., color, size, and position). Based on the calculated Kansei factors, we can determine which regions of a picture will interest viewers. The system calculates this factor for each segmented area and then detects features by comparing each attractiveness value with an attractiveness threshold.

D. Combining Regions

If the distance between two regions is shorter than a distance threshold, the regions are combined into one. Figure 3 (bottom) shows an example of combining two regions that exist within the threshold. If the distance between the two regions is longer than the threshold, the regions are treated separately.

E. Deforming Feature Areas

We used a multi-fisheye technique to treat multi-features [10, 11]. The image was initially mapped onto a mesh and

then deformed by the mesh translation. The system sets multi-viewpoints for zooming on the basis of the desired features. The system calculates the inside of the feature areas linearly and the outside of them non-linearly.

Figure 4 shows simple example of multi-deformation. In Figure 4 (1), there is no borderline to deform multi-features. In Figure 4 (2), the system sets a borderline between two features and deforms each feature. The system also treats features as a kind of node, so it can move the node position freely. In Figure 4 (3), the system first moves features merely by controlling the mesh’s UV coordinates (texture mapping parameter) and then deforms them by using the mesh’s XY coordinates (position parameter). By the fisheye method, the structure of the mesh is translated based on the relationships between the center of each feature area and a constant force that controls deformation size. We can thus control the size of a feature area by adjusting the force interactively.

With our method, only the feature areas were deformed and were made larger and clearer than in normal images, while the non-feature areas became smaller than in normal images. Thus, feature and non-feature areas can be distinguished more clearly even if an image has multi-features. In addition, the two processes used (identifying features and deforming them) were very simple and were based on almost processing speed for a particular image (256 pixels x 256 pixels) and base mesh (200 polygons) using a normal PC. Since we can adjust the image resolution and mesh flexibly, we can adapt them to the capabilities of the target device.

F. Results

Two groups of thumbnails are shown in Fig. 5 (121 images, each 256 x 256 pixels). The original images are shown on the left and the images deformed using the image enforme technique are shown on the right. Although the feature areas in the original images were found using image processing, the characteristic objects (human faces, buildings, symbols, etc.) were successfully identified and deformed. Since most of the original images contained such objects, the subjects on the right are much clearer.

Figure 6 shows some examples of effective and ineffective deformations using the contents enforme technique. When the feature areas were small, such as those shown in (1) and (2), the objects were deformed more effectively. Although the object in (2) was initially difficult to identify due to its small size, the deformation clarified the image. When there were multiple small areas, such as those shown in (3) and (4), the deformation again clarified the objects in the image. Two groups of thumbnails are shown in Fig. 7. Those on the left are 16 normal thumbnails and those on the right are 25 deformed thumbnails. Although the deformed images are smaller, the feature areas are about the same size as in the originals. Two pairs of representative thumbnails are shown in Fig. 8. The objects in the feature areas, a building and two people, are virtually the same size as in the originals. Our technique thus produces images suitable for display on small devices.

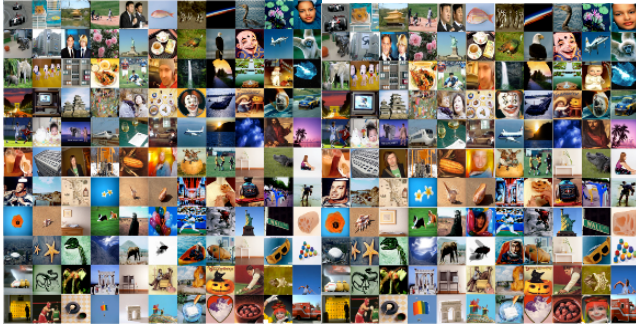


Figure 5. Example thumbnails deformed using our technique. Original thumbnails are shown at left and deformed thumbnails are shown at right.



Figure 6. Comparisons of normal images (left) and images deformed using image enforme technique (right).



Figure 7. Comparison of thumbnails of two different sizes. Originals are shown at left and images deformed by using image-enforme technique are shown at right.



Figure 8. Representative thumbnails taken from images in Fig. 8. Although the deformed images are smaller, the feature areas are about the same size as in the originals.

#### IV. APPLICATIONS

##### A. Net-meeting

We think our deformation technique should be useful for net-meeting systems by making contents more fun and interesting. Net-meeting systems operate using an input image from a camera. The “attendees” are linked by a network and communicate “face to face” (Fig. 9 (top-right)). Net-meeting system users communicate with other users

through the camera image and their voice, so users’ facial expressions are one of the most important elements. However, conventional net-meeting systems focus on how to achieve an effective network model or how to share information, rather than on interaction techniques. With our deformation technique, net-meeting content could be enhanced. For example, our method recognizes a user’s face as a feature area and deforms it automatically to enlarge it (Fig. 9 (bottom, left to right)). After deforming a face area, the user’s face becomes slightly comical (big face and small body). Thus, the deformation can make net-meetings more fun and interesting. Moreover, as our system can treat multi-features, more than two face areas are deformed in the same frame. This kind of net-meeting system is suitable for casual use cases, such as communicating with relatives and friends or playing video games.



Figure 9. Net-meeting enhancement with contents-enforme technique. In this case, a user’s face is deformed and made bigger.

Also, user expressions are enhanced, which should improve communication and clarify who is speaking. In conventional meeting systems, all attendees’ faces are the same size, so voice is the only trigger to understand who is speaking. However, by integrating voice information with our deformation, users can clearly understand who the speaker is. Thus, the deformation method is also useful for regular meetings such as business meetings.

##### B. Video Contents

The use of video contents, such as recorded TV programs and home videos, is becoming widespread in everyday life, and many people have to handle large amounts of contents. Thus, it is important for users to understand video contents quickly. Generally, to find out what the video contents are people have to play them one by one manually at high speed. In this case, deformation is an effective way to roughly determine the nature of the contents. There are two different types of deformation for treating video contents: frame deformation and timeline deformation.

As video contents consist of a number of frames set on a time axis, they are based on two elements, the frame data and the time data. Both elements have feature areas, so our technique can be used for video deformation. Since the frame data is deformed in the same way as net-meeting data, we here describe only deformation of the time data. Our

technique can also treat the time data along with the content features and deform the video contents. We focus here on using it to view or compress the contents. The deformation is done by controlling the video play speed on the basis of the frame features and by simple image processing that combines non-linear and linear deformation.

A simple example of timeline deformation is shown in Fig. 10. This example focuses on a soccer game. The actions are reflected by the differences between each frame and the average frame. The histogram at the top shows the differences between each frame and the average frame. The average frame is created first from all the frame data; each frame is then compared with the average one. In a soccer game, one camera catches the overall image of the field and many other cameras closer to the field catch more detailed images. The average frame is the one captured by the main camera. When an exciting moment occurs, such as a free kick or a chance for a goal, other cameras zoom in on the action. The frame data thus dynamically changes and these changes are obtained by comparing the frame caught by each camera with the average frame.

The feature areas are set by controlling the threshold level (Fig. 11 (1)). The play speed is calculated by the fisheye algorithm. Thus, the play speed depends on the distance between an area and a feature area (Fig. 11 (2)). If an area is close to a feature area, the area is played at almost normal speed. On the other hand, if an area is far from a feature area, the area is played at quicker speed. The feature area contents and close areas to the feature are played at normal speed and the non-feature area contents are played quickly. As a result, the total playing time can be as little as one-fourth that of the original video (Fig. 11 (3)).

V. DISCUSSION

A demonstration of our contents enforme technique during a laboratory open house prompted many interesting reactions and comments from the visitors. Although this was not a scientific evaluation, it still provided useful input.

A. Images Enforme and Thumbnails

The visitors could generally identify the deformed images more clearly than the original images. Most visitors could understand the concept of providing partly deformed images for use as thumbnails and the value of retaining reduced non-feature areas. They quickly comprehended the efficiency of doing this by comparing normal and deformed thumbnails (Fig. 5) and different size thumbnails (Fig. 7). People often take several similar pictures with a digital camera, so the zoom capability is useful for browsing among such images. The visitors could easily identify the objects in the deformed image even when it was difficult to identify them in the normal image (Fig. 6). Also, since the features are produced through a combination of non-linear and linear deformation, most of the deformed images, especially the faces, looked quite funny. Thus, these deformed images are useful for comic creation [15].

We think our approach offers two advantages for information retrieval compared to conventional approaches.

One is that it supports memory-oriented browsing. Generally, users can remember the most impressive part of the pictures they took with a camera or created using paint tools. The elements often make a strong impression on them and become a trigger for browsing. Although our approach deforms only part of an object, we think it is particularly effective in this regard. The other advantage is that it supports more effective browsing among similar images because the total information is kept the same as for the original image by combining non-linear and linear deformations. Also, if two images are similar, the differences between the deformed images are greater than the differences between the original images. Thus, our approach can enhance differences between similar images.

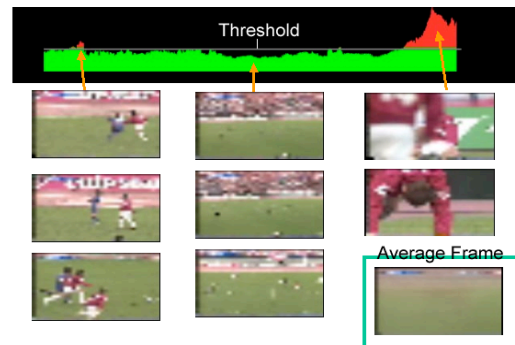


Figure 10. Retrieving feature data. The histogram shows the difference between a frame and the average frame.

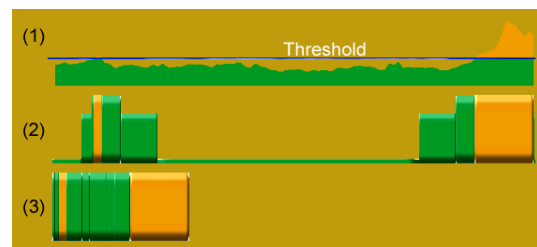


Figure 11. Deformation of video data: the video data is deformed along with the feature area. The feature areas are played at normal speed and the non-feature areas are played more quickly.

Many visitors asked questions about the image enforme calculation and the error rate. Although they could easily imagine it working effectively in small devices, they imagined that much computational power would be needed to deform images. Actually, the image segmentation and fisheye calculation require only moderate computational power. However, since we can adjust the image resolution and mesh flexibly, we can adapt them to the capabilities of the target device. The error rate is related to the size of the feature area and is important because a deformed image is not effective if the original features cannot be retrieved precisely. Since our deformation is based on fisheye calculation, objects close to deformed features are also deformed. Moreover, both the feature and non-feature areas still reside in the deformed image without correct features. In contrast, with cropping, unrecognized feature areas are cropped.

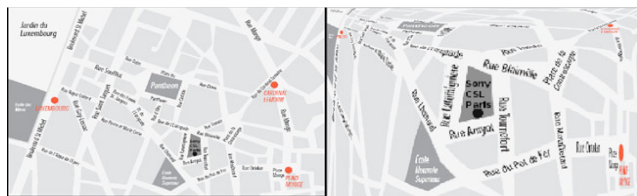


Figure 12. Deformed map: A simple example of a deformed map created by combining the contents-enforme technique with optical character recognition.

Users also suggested many possible applications such as deformed CD jackets and maps. Thumbnails of deformed CD jackets could be used in applications for browsing the contents of music storage devices. Moreover, maps could be deformed to give prominence to areas of interest. A simple example of a deformed map is shown in Figure 12.

### B. Video Enforme

Our concept for deformed video contents was well received. Many deformation techniques for video contents are available, especially in the product field [12]. Although they also focus on the feature areas, they create a digest by collecting only the feature areas. While users can quickly view the contents, they can understand only parts of the contents, so they may have trouble understanding the relationships between the whole story and the features. They are thus quite similar to image cropping in that the non-feature areas are discarded. In contrast, our technique focuses on the non-feature areas and maintains the relationships between them and the feature areas. The contents in non-feature areas are played more quickly, enabling users to view the entire video and understand the importance of the contents based on the playback speed. By adjusting the threshold level, a user can set the total playtime.

It is difficult to define the features of TV programs by using one universal method because TV programs cover a very wide range of areas. We demonstrated different types of feature retrieving for three different types of TV programs. We can retrieve some information about a TV program through the Web or TV listings. Sound and face detection are again useful features for deformation. When an exciting moment occurs, such as the chance for a home run or a goal, the announcer’s voice generally becomes louder. Thus, a part where the voices are louder often contains an interesting or emotional scene, so voice-based deformation is promising for use with video contents. Moreover, scenes with the user’s favorite actor, i.e., someone previously registered to the system by the user, would play at normal speed.

Our deformation focused on the time axis can be used effectively for sound deformation in, for example, a song. Most sound data contains refrain or chorus parts [13]. These parts often include the “hook” of a song and are a kind of feature area. It should be possible to deform sound data based on such features and control the playing speed accordingly. A sound digest should be an effective way to browse music available through the Web because it focuses on feature areas and plays more quickly than the original sound data.

## VI. CONCLUSION

We have described our contents enforme technique for deforming and reducing the features of contents (image and video). The technique uses image processing (face detection and image segmentation) and zooming (non-linear and linear). We showed some examples of our deformation and discussed its effectiveness. Since this technique has potential applications in various areas, including video content browsing and net-meeting systems, we mentioned these applications in addition to describing our concepts.

We are planning to improve our technique by combining it with conventional information visualization systems, which focuses on the information layout. Combining our approach with VelvePath [14]

## REFERENCES

- [1] S. H. Ling, B. B. Bederson, and D. W. Jacobs, Automatic Thumbnail Cropping and its Effectiveness, *In Proceedings of ACM UIST 2003*, pp. 95–104, 2003.
- [2] X. Fan, X. Xie, W. Ma, H. Zhang, and H. Zhou, Visual Attention Based Image Browsing on Mobile Devices, *In Proceedings of ACM Multimedia 2003*, pp. 148–155, 2003.
- [3] V. Setlur, S. Takagi, M. Gleicher, and B. Gooch, Automatic Image Retargeting, *In Conference Abstracts and Applications of ACM Siggraph 2004*, pp. 4, 2004.
- [4] F. Liu, and M. Gleicher, Automatic Image Retargeting with Fisheye View Warping, *In Proceedings of ACM UIST 2005*, pp. 153–162, 2005.
- [5] G. W. Furnas, Generalized fisheye views. *In Proceedings of the ACM Tran. on Computer-Human Interaction 1, 2*, pp. 126–160, 1994.
- [6] K. Sabe and K. Hidai. Real-time Multi-view Face Detection using Pixel Difference Feature, Recognition, *In proceedings of SSII 2004*, 2004.
- [7] R. Nock and F. Nielssen. Statistical Region Merging, *Transactions on Pattern Analysis and Machine Intelligence (TPAMI) IEEE CS Press 4*, pp. 557–560, 2004.
- [8] S. Tanaka, S. Inoue, M. Ishikawa, and S. Inoue. A Method for Extracting and Analyzing “Kansei” Factors from Pictures, *In Proceedings of IEEE Workshop on Multimedia Signal Processing 1997*, pp. 251–256, 1997.
- [9] S. Tanaka, A. Planete, and S. Inoue. A Foreground-Background Segmentation Based on Attractiveness. *In Proceedings of CGIM 1998*, pp. 191–194, 1998.
- [10] M. Sarkar, S. S. Snibble, O. J. Tversky, and S. P. Reiss. Stretching the Rubber Sheet: A Metaphor for Viewing Large Layouts on Small Screens, *In Proceedings of ACM UIST 1993*, pp. 81–91, 1993.
- [11] L. Bartram, A. Ho, J. Dill and F. Henigman. The Continuous Zoom: A Constrained Fisheye Technique for Viewing and Navigating Large Information, *In Proceedings of ACM UIST 1995*, pp. 207–215, 1995.
- [12] Iitokomi  
<http://prius.hitachi.co.jp/go/prius/pc/2005sep/iitokomi/index.html>  
[retrived: December, 2011]
- [13] M. Goto. SmartMusicKIOSK: Music Listening Station with Chorus-Search Function, *In Proceedings of ACM UIST 2003*, pp. 31–40, 2003.
- [14] H. Tobita. VelvetPath: A Layout Design System with Sketch Manipulations, *In Proceedings of EuroGraphics 2003 Short Presentation*, pp. 137-144, 2003.
- [15] H. Tobita, K. Shibasaki. EnforManga: Interactive Comic Creation Using Drag-and-Drop and Deformation, *In Proceedings of IEEE Multimedia 2009*, pp. 269-274, 2009.
- [16] H. Tobita, and F. Nielsen, Image Enforme: Automatic Deformation of Image for Multi-features without Information Loss, *Pervasive 2009 (late breaking result)*, 2009.



# Authenticated Tangible Interaction using RFID and Depth-Sensing Cameras

## Supporting Collaboration on Interactive Tabletops

Florian Klompaker, Holger Fischer, Helge Jung

C-LAB

University of Paderborn

Paderborn, Germany

e-mail: florian.klompaker@c-lab.de, holger.fischer@c-lab.de, helge.jung@c-lab.de

**Abstract**—Interactive large screen displays like tabletops or walls can enhance the interaction between humans and computers. A major topic is the collaboration between multiple simultaneous interacting people. However, most systems suffer from the problem that a distinction of different users is not possible. Hence, in this paper the authors present a work in progress approach of combining various existing technologies in order to enable personalized authenticated tangible interactions on a tabletop. Therefore, authentication using Radio-Frequency Identification in combination with depth-sensing cameras is used. We demonstrate the feasibility of the approach, the multiple advantages for interaction and give an outlook on further activities and lessons learned.

**Keywords**—*natural user interfaces; multitouch interaction; tangible interaction; interactive table; authentication*

### I. INTRODUCTION

Nowadays large multitouch displays like walls and tabletops can be found in a variety of use cases. Researchers have discussed the advantages like the direct interaction paradigm for a long time (e.g., in [1]). Because of their size, such displays are frequently related to multiuser and collaborative scenarios. It has been shown that large multitouch displays and especially tabletops are able to increase productivity and support the process of collaborative decision-making [2][3].

Additionally, in current research, many projects have identified the beneficial use of ‘tangibles’ (arbitrary physical objects which can act as input devices for various tasks). Tangibles that are placed on top of an interactive tabletop can be useful especially while performing complex tasks with even more complex tools, like a Computer-Aided Design (CAD) application [4]. Ishii and Ulmer have introduced tangible user interfaces to a broader community in 1997 [5]. They describe tangibles as physical objects that are graspable and act as computational input and output device at the same time. It has been shown that such kind of Human Computer Interaction (HCI) can improve working processes considerably. This is mainly achieved by giving physical form to digital information and by employing physical artifacts both as representation and as controls for computational media [1]. Associating input and output within the same interface improves the awareness and makes input devices more natural and intuitive since real world tools can be copied in shape and functionality.

However, multitouch displays and tabletops, that are able to detect and track fingers, as well as tangibles suffer from one disadvantage: They are not able to analyze ‘who’ is currently touching and interacting. However, user identification is desirable in collaborative working processes. It enables user authentication and, therefore, it is possible to apply different rights to different users. Further on, for researchers who are working on HCI analysis on tabletop devices, it is of particular interest to track interactions along with information about the users that caused it.

Therefore, in this paper, we present a work in progress approach that uses Radio-Frequency Identification (RFID) technology and depth-sensing cameras in order to allow authenticated touch interactions and authenticated tangible interaction on a tabletop device.

We will first present related work before we introduce our approach. We will then discuss some lessons learned before we sum up.

### II. RELATED WORK

In this section, we present relevant related work that focuses on advantages of multitouch and tangible interaction for collaborative scenarios as well as on technical approaches for user authentication.

Considering topics like security and rights management personalized inputs and actions are desirable. Also, personal reasons are important since nobody likes to enter private data on a public visible virtual keyboard while anybody could be watching. Current systems are not able to distinguish between the authenticated user and the other users working on an interactive tabletop except for some technically complex installations.

The DiamondTouch approach from Mitsubishi Electric Research Laboratories is a multi-user tabletop device [6]. According to a collaborative work scenario the users can interact simultaneously without interfering each other. The surface, on which the image is projected, consists of a special layer with insulated antennas. The touches are transmitted via signals through the antennas. These signals are capacitively coupled through the user and her/his chair to a receiver that identifies the areas of the table that are touched by her/him. A connected computer identifies the user and her/his touches and is able to interact based on this information.

The disadvantage of this approach is that only single touches are recognized by the system. In addition the users have to sit on a chair all the time. Hence, interacting around a tabletop is not really possible with this approach. The DiamondTouch can indeed distinguish between multiple users, but it does not enable a concrete identification or authentication. Further on tangibles cannot be detected by the device.

Some systems for people or item tracking are available on the market. However, the implementation of such systems involves extraordinary effort including setup and calibration. Regarding optical tracking some special patterns have to be attached to tracked objects (e.g., ARTrack [7]). Related to this, the users have to interact in a kind of artificial surrounding. Nor do such approaches allow authentication.

Regarding the tracking of tangibles, most tabletop devices use fiducial markers on the bottom of the object that are recognized by a camera. There are other technologies to detect objects in space (like the already mentioned ARTrack), not even just on a surface. Tracking technologies from the field of Virtual Reality for example are quite expensive, difficult to set up and mostly unmovable. There are approaches that introduce partially transparent patterns and therewith allow the stacking of tangibles [8][9]. Since stacking is recognized by the system more advanced interaction techniques on interactive tabletops are possible. Unfortunately here the use of opaque tangibles is not possible and the tangibles have to fit perfectly on top of each other requiring a very precise handling. The Gecko TUI uses magnets to create unique patterns [10]. These tangibles offer advanced interaction methods like detecting touch regions on the tangibles itself. It is shown that all these TUIs for advanced interaction require some kind of inner electronic circuit and/or mechanic making them less flexible and complicate to use.

Therefore we created the framework dSensingNI (**Depth Sensing Natural Interaction**) [11]. It detects tangibles on a tabletop device and is able to analyze complex interactions like stacking, grabbing&moving and releasing. dSensingNI uses a depth-sensing camera and intelligent occlusion aware algorithms for image analysis. But, like all mentioned approaches, dSensingNI is not able to identify specific users and enable authentication.

Due to the missing authentication possibilities of existing systems, we present an approach that uses RFID technology as well as depth-sensing cameras to realize authenticated user interaction on an interactive tabletop.

RFID chips have been developed to track deliveries of goods. The chips can be either integrated in cards, stickers or directly in products. They are useful for the identification of items and also for the identification of their owner in case of an ID card. According to their production in bulk, RFID chips are a cheap possibility for a sophisticated kind of authentication, which is also conceivable for multitouch tables. Nevertheless RFID in combination with interactive tabletops are only mentioned in approaches concerning the identification of products and their presentation [12].

### III. APPROACH

Our general idea is to combine OpenNI [13] based skeleton tracking using a depth-sensing camera with an RFID reader. After associating a read RFID tag to a user profile our vision algorithm tries to detect a skeleton's hand near the location of the RFID reader that is mounted beneath the table's surface (see Figure 2). Therefore, the algorithm is able to assign an RFID tag to a detected skeleton. The hand movements, touch events and tangible interactions of that user are 'authenticated' from then on. In the next subsection we will explain the mapping of touch events to users and afterwards the combination of skeleton tracking with dSensingNI for authenticated object interactions.

#### A. Authenticated Touches

A sketch of the view from the depth-sensing camera Microsoft Kinect on our local tabletop (called useTable [14]) is shown in Figure 1 (left). The task is to correct the perspective distortion caused by the camera not being located right above the center of the surface. The desired result is a mapping on the screen like shown in Figure 1 (right). Static configuration data provides the three-dimensional location of the four corners of the projection surface. From this information, a geometric plane E can be constructed by using one of the corners (e.g., TL) and vectors to the neighboring corners as already shown in Figure 1; d can be calculated by testing with known points of the plane (TL, TR, BL or BR).

$$E : TL + r \cdot \vec{g} + s \cdot \vec{h} = d \quad (1)$$

The position of a hand as detected by the skeleton tracking framework is referred to point P – which is now being projected onto the plane E. To do this, the normal vector  $n = g \times h$  of E is used to intersect a line from P into the plane along n:

$$\vec{n} \cdot P - d = P + t \cdot -\vec{n} \quad (2)$$

By solving for t and calculating with the real values, we get PT – the projection of P into the plane E as if a shadow would be casted by the light coming from above the projection surface. By solving for r and s in equation 1 we get the logical position of projected point P on the plane - both margins need to be between 0 and 1 to be on the plane. By applying these factors to the real screen we can map the input from the skeleton tracking to a position on the actual projection.

The algorithm is very fast as the involved mathematics is not complex and the necessary computations can be done in only a few cpu instructions. The dominating factor regarding performance remains the tracking of the skeleton itself and, thus, the achievable framerate is only determined by the OpenNI tracking. In our tests, we could use the maximum framerate delivered by OpenNI in real-time without having to drop any frame. One optimization not even implemented by us is that some of the projection

calculations can be pre-computed before the capturing is started.

The algorithm presented above is applied on the incoming data to map hands that were detected by the skeleton tracking in the depth-image onto the actual presentation surface of the tabletop. This information is shown by the prototype in a separate window (see Figure 2 for an example screenshot).

While doing the visual tracking, the prototype uses an RFID reader to detect tags. For this task, the libtagReader embedded in the open-source project tageventor [15] is being used to periodically poll a list of present tags. This allows a fluid change of users since new ones can join the group at any time. When the visual tracking detects a hand in the area of the RFID reader, the software checks if there are tags present on the reader. All the tags are then assigned to the corresponding skeleton. The person is continuously tracked by the visual system until the person gets out of view. In that case she/he has to identify herself/himself again.

Currently, the prototype shows the assigned RFID tag identifiers alongside the tracked hands, but doesn't combine the multitouch input capability with the detected touch data from the useTable. We use the touch data from the table directly to track more precise touch events. Even though the depth-sensing camera can detect or estimate touch points on it's own [16]. Therefore, an interactive table is optional in this setup.

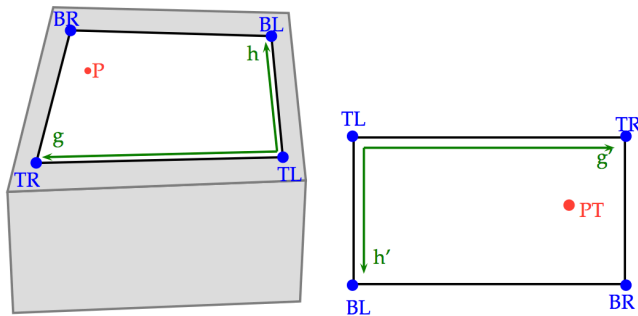


Figure 1: Symbolic representation of the projection surface as seen by the Kinect (left); Desirable target representation of a hand hovering over the projection surface (right)

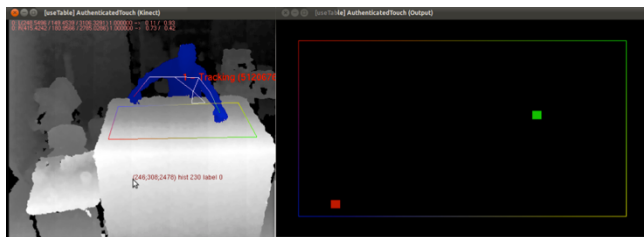


Figure 2: Authenticated touches on a tabletop device. Depth image with skeleton tracking activated (left) and detected finger touches that are assigned to two different users (right).

From the above, it is clear that – upon a multitouch event – the application can check if an authenticated user

was detected in the relevant area of the projection surface and react upon this. A possible reaction could be to ignore an unauthenticated touch or allow special interaction only for selected persons.

### B. Authenticated Objects

After realizing and testing the authenticated touch interaction we used the dSensingNI approach to detect objects on the useTable and their relation to each other [11]. dSensingNI is able to detect and track hands, objects and complex gestures. Objects are identified using their size and volume that can be analyzed from a depth image. Knowing which user is currently grabbing an object (using the information from the authenticated touch approach) allows the system to create user-object relationships and therefore record the user's interaction and assign personal objects to the users. In order to be able to detect even smaller objects we had to use a second depth-sensing camera (here again we used a Microsoft Kinect) as well as a second tracking computer. Since dSensingNI is able to transmit all detected object data using the TUIO protocol [17] the main application is easily extendable.

Since both our cameras for skeleton and for tangibles tracking use projected infrared patterns, they interfere each other in principle. Therefore we had to create a setup to avoid these interferences by using a top down setup for the tangibles tracking with dSensingNI and a front perspective for the skeleton tracking for user identification. Figure 3 (left) shows dSensingNI with a detected hand that just grabbed one out of three objects that are lying on the useTable. Figure 3 (right) shows a pointing gesture towards a tangible.

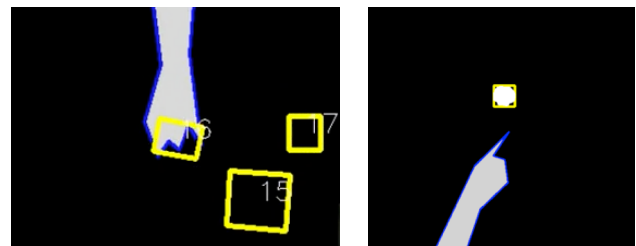


Figure 3: dSensingNI is used to detect objects, hands and grabbing gestures (left) as well as pointing gestures (right)

Besides detecting interactions with tangibles, dSensingNI is also able to detect freehand pointing gestures as shown in Figure 3 (right). This enables for example the selection of unreachable objects on an interactive tabletop. Using our combined approach this kind of selection is now also assigned to a specific user profile and therefore 'authenticated'.

## IV. DISCUSSION

We implemented a simple map based application as a first prove of concept (see Figure 4). Here, we are able to apply authenticated touch so that only one predefined user is

allowed to manipulate the map (zooming and moving). Further on objects on the useTable are detected and assigned to the user who put it on the table or who performed the last interaction with the object.

In this setup, we had a few interference problems with the two depth-sensing cameras. In order to keep the skeletons in view we had to move the skeleton tracking camera a little towards the ceiling.

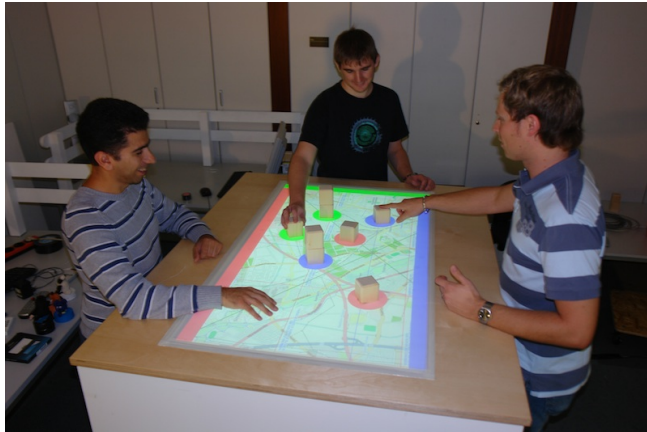


Figure 4: Authenticated collaborative tangible interaction on a tabletop

Compared with other solutions that recognize interaction in space, we presented a cost-effective approach that is also easy to set up. Even though our solution is not suitable for high-risk applications yet, because user IDs can be taken over by anyone in front of the depth-sensing camera quite easily, it demonstrates possibilities and scenarios for authenticated interactions on tabletops quite well.

Thinking of a tabletop system for collaborative mission planning in the field of disaster control management like that one we presented in [18], some workers might have additional rights in decision making others don't have. According to our approach users can authenticate themselves to the system and a specified set of interactions can be activated. Users can use tangibles or multitouch in compliance with her/his position in the organizational hierarchy.

## V. CONCLUSION

In this paper, we presented a work in progress approach to enable personalized authenticated tangible interactions on an interactive tabletop device. Therefore, we used our dSensingNI framework that detects tangible objects on the surface of tabletops and evaluates complex interactions like stacking or grouping. In addition we extended this approach using RFID technology and a second depth-sensing camera for skeleton tracking. As a result, our prototype enables authenticated multitouch and tangible interaction for multiple users interacting on a tabletop. The interaction of multiple simultaneous interacting people can be traced and their steps in a solution finding process can be reviewed, e.g., in a HCI evaluation.

## REFERENCES

- [1] L. Yu, P. Svetachov, P. Isenberg, M.H. Everts, and T. Isenberg, "FI3D: direct-touch interaction for the exploration of 3D scientific visualization spaces". In: IEEE Transactions on Visualization and Computer Graphics, 16(6), 2010, pp. 1613-1622.
- [2] E. Hornecker, P. Marshall, N.S. Dalton, and Z. Rogers, "Collaboration and interference: awareness with mice or touch input". In: Proc. Of the conference on Computer supported cooperative work, pp. 2008, pp. 167-176.
- [3] V. Ha, K.M. Inkpen, T. Whalen, and R.L. Mandryk, "Direct Intentions: The Effects of Input Devices on Collaboration around a Tabletop Display". In: Proc. of IEEE International Workshop on Horizontal Interactive Human-Computer Systems, IEEE, 2008, pp. 177-184.
- [4] R. Aish, "3D input for CAAD systems". In: Computer Aided Design, 11(2), 1979, pp. 66-70.
- [5] H. Ishii and B. Ullmer, "Tangible Bits. Towards seamless interfaces between people, bits and atoms". In: Proc. of the Conference on Human Factors in Computing Systems, 1997, pp. 234-241.
- [6] P. Dietz and D. Leigh, "DiamondTouch: A Multi-User Touch Technology". In: Proc. of UIST 2001, the 14<sup>th</sup> Annual ACM Symposium on User Interface Software and Technology, 2001, pp. 219-226.
- [7] <http://www.ar-tracking.de>, 15.11.2011.
- [8] T. Bartindale and C. Harrison, "Stacks on the surface: resolving physical order using fiducial markers with structured transparency". In: Proc. Of the International Conference on Interactive Tabletops and Surfaces, 2009, pp. 57-60.
- [9] P. Baudisch, T. Becker, and F. Rudeck, "Lumino: tangible blocks for tabletop computers based on glass fiber bundles". In: Proc. Of the 28th international conference on Human factors in computing systems, 2010, pp. 1165-1174.
- [10] J. Leitner and M. Haller, "Geckos: Combining Magnets and Pressure Images to Enable New Tangible-object Design and Interaction". In: Proc. of the Conference on Human Factors in Computing Systems, 2011, pp. 2985-2994.
- [11] F. Klompaker, K. Nebe, and A. Fast, "dSensingNI – A Framework for Advanced Tangible Interaction using a Depth Camera". In: Proc. of the Sixth International Conference on Tangible, Embedded and Embodied Interaction (TEI), 2012, in press.
- [12] Y. Huang, S. Wang, and Z. Deng, "An Interactive RFID-based Multi-touch Product Display System". In: Proc. of the International Multi-Conference on Complexity, Informatics and Cybernetics, 2010.
- [13] <http://www.openni.org>, 15.11.2011.
- [14] <http://www.usetable.de>, 15.11.2011.
- [15] <http://code.google.com/p/tageventor/>, 15.11.2011.
- [16] A.D. Wilson, "Using a depth camera as a touch sensor". In: Proc. Of the ACM International Conference on Interactive Tabletops and Surfaces, 2010, pp. 69-72.
- [17] M. Kaltenbrunner, T. Bovermann, R. Bencina, and E. Costanza, "TUIO - A Protocol for Table-Top Tangible User Interfaces". In: Proc. of the 6th International Workshop on Gesture in Human-Computer Interaction and Simulation, 2005.
- [18] K. Nebe, F. Klompaker, H. Jung, and H. Fischer, "Exploiting New Interaction Techniques for Disaster Control Management using Multitouch-, Tangible- and Pen-based Interactions". In: Proc. of the 14th International Conference on Human-Computer Interaction, LNCS 6776, Vol. 16, 2011.

# Generic Brain-computer Interface for Social and Human-computer Interaction

Julita de la Vega Arias, Christoph Hintermüller, Christoph Guger

g.tec medical engineering GmbH and Guger Technologies OG  
 Schiedlberg, Austria  
 delavega@gtec.at, hintermueller@gtec.at, guger@gtec.at

**Abstract**—After suffering a more severe disease like spinal cord injury or stroke patients are often not able to interact or even communicate with their environment anymore, especially at the beginning of rehabilitation. Brain-Computer Interfaces (BCIs) can substitute this temporarily lost communication channels and might support rehabilitation by providing an alternative way for controlling a computer only by thoughts without any muscle activity. This enables the patient to communicate by writing letters on the screen, to stay socially in contact with friends or people outside the rehabilitation facility by participating in games like *Second Life*, where they may appear as healthy persons. Another application is to control items in their room connected to the BCI system like the lights which can be turned off and on as it can be done in a virtual smart home without leaving the bed. In this paper, the technology of such BCIs and the mentioned applications are described utilizing the P300 approach. A generic BCI interface is presented, which allows controlling them and concurrently and transparently switching among them. The results of a recent study show that a BCI can be used by patients suffering from cervical spinal cord injury almost as well as by healthy people which encourages us to think it may assist rehabilitation regarding the social aspect. A variety of BCI can be implemented with the aim of provide them a more active and accessible lives.

**Keywords**—Brain-Computer Interface; BCI; P300; Visual Evoked Potentials; speller; *Second Life*; *Virtual Smart Home*

## I. INTRODUCTION

Many disorders can affect or even completely damage the usual communication channels a person needs to communicate and interact with his or her environment. Spinal cord injury, stroke and Amyotrophic Lateral Sclerosis (ALS), for example, can result in partial or complete loss of voluntary muscle activities including speech. In such severe cases, a Brain-Computer Interface (BCI) might be the only remaining possibility to communicate [1]. BCI based on Electroencephalography (EEG) provides a new non-invasive communication channel between the human brain and a computer without using any muscle activity. It measures and analyzes the electrical brain activity during predefined mental tasks and translates them into the corresponding action intended by the user. But, even for less severe levels of affection, a BCI can improve quality of life of partially paralyzed patients during rehabilitation by enabling them to control a computer and specially prepared electronic devices, or to stay in contact with friends through social networks and games, for example.

BCIs based on P300, positive-going waves with a peak typically at typical latency of 300-1000 ms from stimulus

onset, provide goal-oriented control and are mainly used for spelling devices. Farwell and Donchin were one of the first pioneers who used the P300 for communication [2]. Furthermore, this approach can also be used for game control [3] or navigation (e.g., moving a computer mouse [4]). The P300 evoked potential is elicited when a unlikely event occurs randomly between events with high probability. It manifests itself in a positive deflection in the amplitude of the EEG signal around 300ms after a visual stimulus onset. Different classification techniques have been evaluated for P300 spellers whereby both the Fisher's linear discriminant analysis and the stepwise linear discriminant analysis yielded very robust performances [5]. These classification methods need to be trained on each individual subject to adapt them to the particular brain activity behavior of a person. However, training of such a BCI system can be accomplished within several minutes [6]. Guger et al. demonstrated that more than 70% of a sample population is able to use such a spelling setup with an accuracy of 100% [9].

This paper presents a generic User Datagram Protocol (UDP), Extensible Markup Language (XML) based BCI command interface (according to the recently introduced international BCI interface standard, see [www.gtec.at](http://www.gtec.at)) and its application to implement control interfaces enabling social interactions like spelling, interacting virtually with other participants in *Second Life*, operating Twitter and controlling a virtual smart home. All of these applications are based on the P300 speller principle. Results of a spelling application study done with 100 healthy people will be compared to a recent study also including spinal cord injury patients showing that this type of a BCI would be suited for rehabilitation assistance.

This paper presents different applications (Twitter Interface, Virtual Smart Home Control and a Generic Control Interface ) tested with healthy people. The outcomes from them are shown in different tables in Section III.

## II. MATERIAL AND METHODS

For a P300 spelling device, a 6x6 matrix of different characters and symbols is commonly presented on a computer screen [7]. In single-character mode all characters are flashed in a random order but only one character after each. In row-column mode, a whole row or a whole column flashes at a time. The subject has to concentrate on a specific letter he or she wants to write. The flashing of exactly this character or the corresponding row or column is a relative unlikely event which induces a P300 component in the EEG signal reaching its maximum amplitude around 300 ms after the onset of the flash.

For all other characters, rows or columns no such P300 component is elicited because they are not relevant to the subject currently.

To measure the P300 component acquisition of EEG, signals from 8 electrode positioned mostly over occipital and parietal regions is sufficient [8]. To train the BCI system, an arbitrary word like LUCAS is announced to the system to be aware of which characters the subject is supposed to concentrate on (targets) and which not (non-targets). Each of these letters respectively, each row and column flashes several times e.g., for 100 ms per flash. The subject focuses on each of these letters, one after the other, and increments a mentally running count whenever the letter flashes the subject is currently concentrating on. EEG data of a specific time interval around each flash is then sent to a Linear Discriminant Analysis (LDA) classifier to learn to distinguish the typical EEG signal form of the target characters from the typical signal form of all other non-targets.

The applications described in the following were solely tested with healthy people. All of the following applications are based on the same principle and setup. They basically differ in the content of the matrix where letters may be replaced by symbols or phrases to control the associated applications.

A. P300 Twitter Interface

Through the social network Twitter (Twitter Inc.), users can exchange messages. These messages can be sent over the Twitter website, by smart phones or by SMS (Short Message Service). They are displayed on the author’s profile page and are limited to 140 characters per message. Interfacing possibilities can be extended through the application programming interface provided by Twitter.

The control mask of the classical speller application was extended to add the necessary Twitter control commands in the first two lines (Fig. 1). The remaining characters are used for spelling resulting in a total of 54 possible symbols. Fig. 3 shows the Unified Modeling Language (UML) diagram of required actions for using the Twitter service.

The system was initially trained with 10 target symbols for the BCI user. Then, another user was asking questions through Twitter and the BCI user had to answer each question every other day. A total of 9 questions were asked which means that the BCI user had to use the P300 interface on 9 different days whereas they selected between 6 and 36 characters each day.

Login	Logout	Line	Search	Friends	Post
Inbox	Send	Follow	Leave	Delete	Enter
A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	0	1	2	3
4	5	6	7	8	9
,	.	!	?	⊗	-

Fig. 1 The extended P300 interface control mask for Twitter.

B. Virtual Smart Home Control

A virtual 3D representation of a smart home was designed based on the XVR environment (eXtreme Virtual Reality, University of Pisa) [10] [11]. Seven different control masks were developed for the P300 BCI in order to control the virtual smart home environment.



Fig. 2 Two views of the virtual smart home and the corresponding control masks.

Fig. 2 shows two example views of the virtual apartment with their corresponding control masks. The pictures on the top show a 3D view of the living room (Fig. 2a) and the corresponding main control mask (Fig. 2b) enabling the user to control some of the devices in the living room like the TV set, the room light or the telephone. Fig. 2c shows a bird’s-eye view of the virtual apartment. The living room, for example, can be found in its top left corner and the bathroom is located in the bottom right corner together with the entrance door. Using the related “GoTo” P300 BCI control mask (Fig. 2d), the user can beam himself to 21 different locations in the apartment. Using the seven control masks, it should be possible for users to switch the light on and off, to open and close the doors and windows, to control the TV set, to use the phone, to play music, to operate the video camera at the entrance, to walk around in the house and to move their selfesto a specific place in the apartment.

The subjects were instructed to avoid unnecessary movements and to accomplish some specific tasks like opening the front door, moving to specific places in the apartment or manipulating the light source or the room temperature. Further details on the setup and results can be found in [10].

C. Generic Control Interface

The interface was designed with the intention to control multiple applications and devices simultaneously and to provide a common way of sharing BCI control in a consistent manner among them. The various applications, like the Twitter (Section II.A), differ in the content of the P300 BCI matrix where letters may be replaced by symbols or phrases to control the associated applications.

In a first step, the possible control states and the commands available in each state to move on to the next state are identified using Unified Modeling Language (UML) diagrams, which yield a simple and clear overview of all required elements. Based on the diagrams, the detailed descriptions of the states and the actions to be taken along with the transition

to the next state, an XML file describing the required masks, the position of the different symbols within the mask and the commands to call is generated for each application. Upon startup, the initial interface descriptions of the different applications are merged and symbols are added for switching among them. Fig. 3 shows the UML diagram for the Twitter application. Fig. 1 shows the P300 mask resulting thereof. Whenever a control symbol or character is selected by the user, the BCI emits the command or character string associated with the symbol or executes the related action.

Alternatively, the interface allows for online updating the control mask of the application in reaction to its current internal state or the state of its environment in real-time. Thereby, it sends the adopted description of the control mask to the interface unit through the UDP network socket. At the beginning of the next flash cycle, the BCI interface will load this updated mask and present it to the user for the selection of the appropriate commands.

### III. RESULTS

The interface was implemented for and tested with the two applications, Twitter and Second Life presented above. For the 2 systems, the control states were identified using UML state diagrams. Based on these the control mask, descriptions were generated, merged and the symbols for switching between the applications were added. In the following sections, some the results achieved for each of the applications are discussed.

The EEG data were recorded for all above applications including the interface test applications with a g.MOBilab+ Biosignal Amplification Unit (g.tec medical engineering GmbH, Austria) at 256 Hz sample rate and transferred to the computer wirelessly via Bluetooth®. A MATLAB/Simulink model controls the interface masks, processes the received data via a specific device driver and dispatches the targeted commands via the described UDP XML message passing interface. A notch filter (50 Hz or 60 Hz) and a band pass filter were applied to the signals in order to eliminate possible artifacts before they were down-sampled to 64 Hz. Data from 100 ms before each flash onset to 700ms afterwards were filtered and down-sampled again to get 12 feature values (i.e. samples) per channel and flash. These data chunks were sent to the LDA to determine if a target character flashed or not. The subjects were sitting or standing, dependent on the application, in front of a computer screen and were instructed to relax as much as possible.

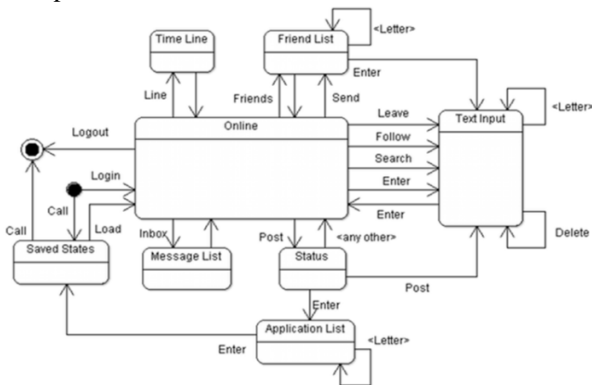


Fig. 3 UML diagram of the actions required to use the Twitter service

### A. P300 Twitter Interface

TABLE I. QUESTIONS AND TEXT INPUT WITH THE P300 TWITTER INTERFACE

Tweets	Number of characters	Duration [hh:mm:ss]	Errors	Errors %	Number of flashes	Time per character [s]
<i>FRIEND: Which kind of Brain-Computer Interface do you use?</i>						
<b>BCI: P300 GTEC BCI</b>	<b>13</b>	<b>00:11:09</b>	<b>3</b>	<b>23,08</b>	<b>8</b>	<b>51</b>
<i>FRIEND: Are you using the g.GAMMAsys?</i>						
<b>BCI: Exactly!</b>	<b>7</b>	<b>00:06:18</b>	<b>1</b>	<b>14,29</b>	<b>8</b>	<b>54</b>
<i>FRIEND: Active or passive electrodes? For explanation: the active system avoids or reduces artefacts and signal noise.</i>						
<b>BCI: Active electrodes</b>	<b>17</b>	<b>00:06:10</b>	<b>0</b>	<b>0,00</b>	<b>5</b>	<b>22</b>
<i>FRIEND: The mounting of the active system is very comfortable. You do not need to prepare the skin first, do you?</i>						
<b>BCI: you are absolutely right</b>	<b>24</b>	<b>00:08:55</b>	<b>1</b>	<b>4,17</b>	<b>5</b>	<b>22</b>
<i>FRIEND: How many electrodes are needed to run the BCI?</i>						
<b>BCI: For P300 we usually use 8 electrodes</b>	<b>36</b>	<b>00:14:21</b>	<b>2</b>	<b>5,56</b>	<b>5</b>	<b>24</b>
<i>FRIEND: What amplifier are you using for the Brain-Computer Interface?</i>						
<b>BCI: g.MOBilab+</b>	<b>10</b>	<b>00:04:42</b>	<b>1</b>	<b>10,00</b>	<b>5</b>	<b>28</b>
<i>FRIEND: h, g.MOBilab+ - this is the g.tec's portable biosignal acquisition and analysis system!</i>						
<i>FRIEND: How long does it take to code the software for the BCI for TWITTER?</i>						
<b>BCI: 3 Weeks</b>	<b>7</b>	<b>00:03:13</b>	<b>1</b>	<b>14,29</b>	<b>4</b>	<b>28</b>
<i>FRIEND: How many characters are you able to write within a minute?</i>						
<b>BCI: 3 To 4</b>	<b>6</b>	<b>00:03:15</b>	<b>0</b>	<b>0,00</b>	<b>5</b>	<b>33</b>
<i>FRIEND: You are using the BCI device for one month? Did you get faster in writing during this period?</i>						
<b>BCI: Yes, from 2 to 4 characters</b>	<b>27</b>	<b>00:06:38</b>	<b>1</b>	<b>3,70</b>	<b>3</b>	<b>15</b>
<i>FRIEND: Great, thank you for answering the questions!</i>						

Table I shows the results of the answers from the BCI user (bold sentences) to the 9 questions (italic sentences). The improvement of the user's performance from the first session to the last one is interesting.

For the first session, it took the user 11:09 min to spell 13 characters with 3 mistakes. The user was instructed to correct any mistake yielding an average of 51 seconds selection time for each character. In contrast to this, the last session was accomplished in a time of 6:38 min for 27 characters with only 1 mistake resulting in an average selection time of 15 seconds per character. The number of flashes per character decreased from 8 to only 3 flashes as well.

**B. Virtual Smart Home Control**

A total of 12 healthy subjects participated in the case study. The system was trained on 42 selected target symbols for each subject. In contrast to the previous applications, the subjects were standing in front of a 3D power wall for projection of the virtual smart home while using this BCI.

In the virtual smart home study, the subjects needed between 3 and 10 flashes per character (mean 5.2 flashes per character) to achieve an accuracy of 95% in single-character mode [10]. Table II lists the accuracy of all subjects for each of the 7 control masks among other parameters. The more symbols a mask contains, the lower is the probability of occurrence of one symbol. This results in an increase of the amplitude of the P300 component in the EEG signal. Also the time needed to select a single character increases because it takes the system more time to flash all the characters.

Interestingly, the subjects achieved only 26.39% accuracy for the “Go to” mask, which is quite bad compared to the other control masks. We think this is due to its different layout. While all the other control masks were arranged in a matrix layout, the “Go to” mask was the only one where the symbols were placed at apparently random positions related to the appearance of the virtual apartment. More detailed results of the virtual smart home study can be found in [10].

TABLE II ACCURACY, NUMBER OF SYMBOLS, OCCURRENCE PROBABILITY PER SYMBOL, NUMBER OF FLASHES PER SYMBOL AND MASK (E.G., 25 X 15 = 375) AND SELECTION TIME PER SYMBOL FOR EACH MASK

Mask	Accu- racy	No. Sym- bols	Proba- bility	No. Flashes	Time per character
Light	65.28%	25	4%	375	33.75 s
Music	76.11%	50	2%	750	67.50 s
Phone	63.89%	30	3.3%	450	40.50 s
Temper- ature	76.39%	38	2.6%	570	51.30 s
TV	65.74%	40	2.5%	600	54.00 s
Move	75.93%	13	7.7%	195	17.55 s
Go to	26.39%	22	4.5%	330	29.70 s

**IV. DISCUSSION**

Extending the standard P300 speller by icons and symbols enables the user to control more complex scenarios. In [10], it is shown that a proper design of interface masks allows control of domotic devices in a virtual smart home study with comparable reliability as it has been reached for the spelling application. This paper discusses the usage of the BCI interface to control smart homes or operating social interaction applications like sending or answering Twitter messages. The generic BCI interface which uses XML based description files simplifies the definition of the control masks. By merging the control masks of several applications like Twitter and environments like a virtual smart home, it is possible to use the different systems concurrently. To switch to another application, it is sufficient to select the symbol or sequences of symbols used to load its control mask. If not running, the application will be started and initialized to receive the control commands from the BCI.

In a recent study, Guger et al. evaluated a P300 speller (6x6 character matrix) in row-column mode with 100 healthy subjects (32 female and 68 male at the age of 27.9 ± 10.9)

using a similar setup as described in this study [9]. Using a subset of 81 subjects, 88.9% of them were able to use the speller with an accuracy level of 80-100% whereas 72.8% reached an accuracy of 100%. The average accuracy level was 91.1% with a spelling time of 28.8 seconds for a single character. Since this speller is based on the same P300 principle, the results indicate that the previously described applications might also perform better using row-column mode instead of single-character mode.

The results of the study show that a P300 brain-computer interface can also be used by patients suffering from more severe diseases. This encourages us to think that a BCI may be helpful during rehabilitation. Especially when people are bound to bed or rehabilitation facilities, it is important to keep them socially engaged. The applications described here help people to keep in touch with friends and other people all around the world through popular social networks and platforms, like Twitter. Another important fact is that patients may appear as healthy people, for example, in Second Life, which may be a benefit regarding psychological aspects to let the disability fade from people’s minds. Furthermore, a BCI can also be used for wheelchair control, which many authors identify as their target application. However, there are still some minimum requirements to operate a BCI. In fact, there is a certain percentage of the population (including healthy people as well) that is not able to operate a specific type of BCI at all, which may have various reasons [13].

**ACKNOWLEDGMENT**

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under BrainAble Project, Grant Agreement nr. 247447. The authors further gratefully acknowledge the funding by the European Commission under Contract FP7-FET-257695 (Virtual Embodiment and Robotic Re-Embodiment).

**REFERENCES**

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767-791, June 2002.
- [2] L. A. Farwell, and E. Donchin, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510-523, December 1988.
- [3] A. Finke, A. Lenhardt, and H. Ritter, “The MindGame: A P300-based brain-computer interface game,” *Neural Networks*, vol. 22, no. 9, pp. 1329-1333, November 2009.
- [4] L. Citi, R. Poli, C. Cinel, and F. Sepulveda, “P300-Based BCI Mouse With Genetically-Optimized Analogue Control,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 1, pp. 51-61, February 2008.
- [5] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, “Toward enhanced P300 speller performance,” *J. Neurosci. Methods*, vol. 167, no. 1, pp. 15-21, January 2008.
- [6] E. Donchin, K. M. Spencer, and R. Wijesinghe, “The mental prosthesis: assessing the speed of a P300-based brain-computer interface,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 174-179, June 2000.
- [7] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, “A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on



- performance,” *Biological Psychology*, vol. 73, no. 3, pp. 242-252, October 2006.
- [8] C. Guger, S. Daban, E. W. Sellers, C. Holzner, G. Krausz, R. Caraballona, F. Gramatica, and G. Edlinger, “How many people are able to control a P300-based brain-computer interface (BCI)?,” *Neuroscience Letters*, vol. 462, no. 1, pp. 94-98, September 2009.
- [9] G. Edlinger, C. Holzner, G. Groenegrass, C. Guger, and M. Slater, “Goal-Oriented Control with Brain-Computer Interface,” in *Proceedings of the 5<sup>th</sup> International Conference on Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*. Berlin, Heidelberg: Springer, 2009, pp. 732-740.
- [10] C. Guger, C. Holzner, C. Groenegrass, G. Edlinger, and M. Slater, “Control of a smart home with a brain-computer interface,” presented at the 4<sup>th</sup> International Brain-Computer Interface Workshop and Training Course, University of Technology Graz, Graz, Austria, 2008.
- [11] F. Nijboer, E. W. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. J. Krusienski, T. M. Vaughan, J. R. Wolpaw, N. Birbaumer, and A. Kübler, “A P300-based brain-computer interface for people with amyotrophic lateral sclerosis,” *Clin. Neurophysiol.*, vol. 119, no. 8, pp. 1909-1916, August 2008.
- [12] B. Allison, T. Luth, D. Valbuena, A. Teymourian, I. Volosyak, and A. Graser, “BCI Demographics: How Many (and What Kinds of) People Can Use an SSVEP BCI?,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 2, pp. 107-116, April 2010.

# A Conversational System to Assist the User when Accessing Web Sources in the Medical Domain

*Marta Gatius, Tsetsegkhand Namsrai*

Software Department  
Technical University of Catalonia  
Barcelona, Spain  
*e-mail: {gatius,tsena}@lsi.upc.edu*

**Abstract**— The aim of our research is to develop a framework of personal conversational assistants, adapted to different types of tasks and users. In particular, we have focused on how a conversational assistant can help the users in several situations related to the medical domain, and thus can help people take care of their health. In this paper we describe how the cultural assistant could help the user in four situations: finding a new specialist, finding the closest pharmacy, consulting a specific drug prescription and making an appointment to see the doctor. Our proposal is based on the integration of language, dialogue and ontologies to assist the user when accessing different types of web sources: informational and transactional services, dictionaries, maps. We have focused in a set of selected scenarios where a conversational assistant can be useful and the conceptual and linguistic resources needed are limited.

**Keywords**-conversational web assistant systems; ontologies; medical domain.

## I. INTRODUCTION

The improvement of language technologies has made possible the development of conversational assistants that help users when accessing the web. The main advantage of language modes of interaction (text and speech) is that they are friendly and easy to use. Natural language modes support different types of web interaction: menus (the system asks the user to choose between a set of options), form filling (the system asks for a very specific information) and commands (the user can express an order, e. g. to perform a transaction). For this reason, the use of language is specially appropriate for devices not allowing easy internet browsing as well as for users who are not used to access the web.

The use of language also facilitates the development of intelligent web systems because language supports phenomena not supported by other modes, such as references to previous entities and descriptions of complex concepts. Besides, speech and text can be integrated with other modes of communication.

However, the use of language also present several disadvantages compared to other modes of communication. On one hand, the usability of conversational assistants (henceforth, CA) is limited because they present more mistakes and misunderstandings than interfaces using other

modes of communication. As a result, tasks efficiency and user satisfaction in CA presents problems to be studied. On the other hand, CA are expensive to develop and difficult to reuse.

In order to solve these problems, most commercial CA restrict the communication to dialogues in which the system asks the user the specific information needed by the application. Those systems can be efficient and easy to use for simple applications: mistakes and misunderstandings are limited because the linguistic and dialogue resources used are adapted to the application. Besides, the cost of the development of the CA is reduced because the language and dialogue resources needed are limited. Those CAs are, however, difficult to reuse.

Limiting language and dialogue resources to the application has proven appropriate for application that require simple data from the user. Efficient and friendly CA for applications that need more complex information involves advanced conceptual, linguistic and dialogues resources.

Information retrieval systems are examples of applications requiring limited language and dialogue resources are. Dialogues needed by those systems are usually simple, as those in the web retrieval system for mobiles described in [1]. Systems performing complex tasks, such as the Medication Adviser (described in [2]) usually require more complex language interfaces, supporting mixed-initiative dialogues and incorporating domain knowledge. The development of those complex language systems still presents several challenges and it is an active research area.

An approach followed in intelligent systems using language to assist the user in complex tasks (or domains) is based on the use of conceptual ontologies. Ontologies facilitate reasoning about the application domain and can also be used to improve communication in different forms: to infer default information, to provide descriptions of domain concepts, to enrich lexical resources, etc.

The use of ontologies also favors the integration and reusability of the different type of knowledge involved in complex language systems: conceptual, linguistic and dialogue. In order to achieve efficient and friendly communication, complex CA are usually developed for a

specific type of application and their adaptation to a different type of application is not easy. The main problem is that the adaptation of the CA to a new application implies the modification of conceptual, linguistic and dialogue knowledge, usually represented in heterogeneous sources. Ontologies representing the application domain can be shared by the system knowledge bases, thus facilitating their integration as well as their adaptation to new applications.

The integration of advanced techniques on language processing and generation, dialogue and knowledge representation is a complex problem that presents several challenges to be solved. There are different approaches to this problem and most research works on the area are focused on a specific aspect of the problem. Examples of relevant systems integrating language, dialogue and ontologies to assist the user in different tasks are the multimodal SmartWeb system [3], the Active platform [4] and several works done in the framework of the TALK project [5].

Smartweb is a multimodal multilingual dialogue system that provides access to a selection of web-based information services (for example, those giving data related to a specific event, such as world football coup). Smartweb is focused on the integration of several communication modes and languages of communication and semantic web techniques.

Active platform is an intelligent system to assist the user in several daily tasks. It is focused on the integration of language, ontologies and advance techniques on artificial intelligence, such as active recognition techniques.

Several prototypes of multilingual dialogue systems using ontologies have also been developed for different applications in the framework of the TALK project. Most of these works are focused on the use of advanced dialogues strategies.

The work we present in this paper is also about the integration of language, dialogue and ontologies to assist the user in several situations, as the relevant works already described. The main difference of our work and those already mentioned is that we have focused on how to assist the users when accessing different types of web sources in a complex domain: the medical domain. For this purpose, we have studied several common situations in which different types of users may find difficulties when accessing web medical sources. We have focused on those situations where using a CA would be useful and the linguistic and conceptual resources required were limited.

Our proposal is based on a previous work on dialogue systems to guide the user when accessing informational and transactional web services (described in [6]). In [6], a prototype was developed that simulated access to two web services of different type: an informational service on cultural events and a transactional service on large objects collection. The results of the evaluation of the prototype showed the usability of the system was high (most of the users had a good impression of the system: 3,58 over 5), but could be improved. In order to improve the system usability, we have studied the use of domain ontologies for fostering the collaborative ability of the system ([7]). In particular, the use of ontologies was studied in two critical situations that

can take place when assisting the user to find information: no results are found and too many results were found (and thus, it is difficult to present them in a clear form). In the first situation, domain ontologies can facilitate the reformulation of the user's query, while in the second situation, the ontology can be used to organize (or summarize) the results.

In this paper, we describe our study on how [6] could be adapted and extended to assist the user in several common scenarios related to the medical domain. The scenarios described are the following: finding a new specialist, finding the closest open pharmacy, consulting a specific drug prescription and stating an appointment to see the doctor. In these scenarios, the user needs to access web sources of different type: informational and transactional services, dictionaries, maps, etc.

In order to study the main challenges when assisting the user in several scenarios in the medical domain, we have implemented a prototype that provides textual access in Spanish to the several web sources selected for each scenario.

The next section gives an overview of the conversational assistant we developed, including a description of the scenarios considered and the prototype implemented. Then, last section draws some conclusions and future work.

## II. OVERVIEW OF THE SYSTEM

This section describes several of the scenarios we have considered for a web assistant in the medical domain as well as the prototype we designed to study how to approach main challenges those scenarios present.

### A. Studied Scenarios

We have studied a selected set of those scenarios, where several users can have some problems when accessing web medical sources and a CA using limited linguistic, dialogue and conceptual knowledge sources could be useful. Those cases of study were used to study how to approach main challenges when assisting the user to access web sources related to the medical domain.

The scenarios described are the following: the user wants to visit a new specialist, find an open pharmacy, consulting a medical prescription and making a reservation for an appointment to a primary health care center. Our main goal when describing these scenarios was presenting the advantages of using a conversational assistant. For this reason, the situations are simplified, in real situations much more choices could be presented to the users.

#### 1) Searching for a Specialist

We have considered the case of a user that pays for a private medical insurance having an informative web page, containing all clinic and hospitals which are covered by this insurance. Most of these pages are filled with drop down menus as that on Figure 1.

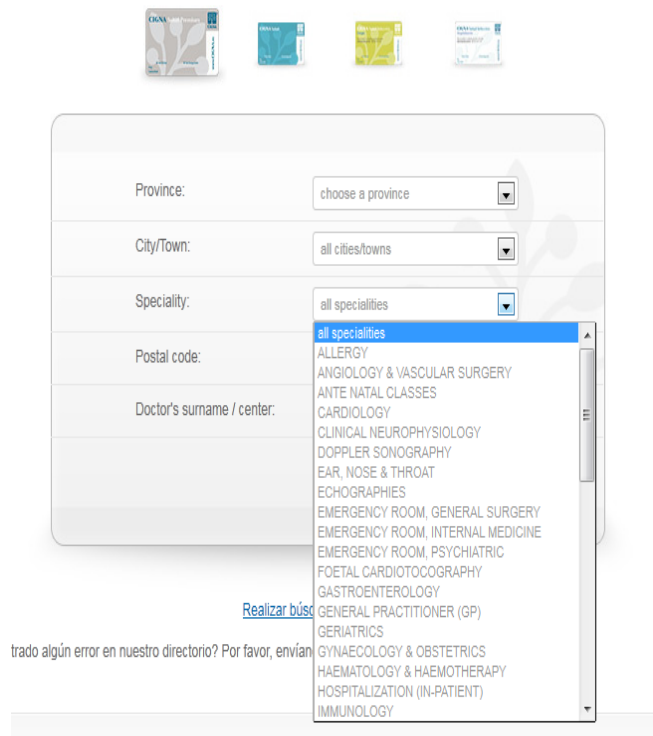


Figure 1. Searching for a Specialist in a web insurance page

TABLE I. EXAMPLE OF DIALOGUE WHEN SEARCHING FOR A SPECIALIST

S1: Hello. This is the automatic health service information system. How can I help you?  
 U1: Hi. I need to find a doctor.  
 S2: Which specialist do you need? You can say the common name.  
 U2: Ear doctor please.  
 S3: Would you like to search in a big hospital or would you prefer to search in smaller equipments?  
 U3: A big hospital.  
 S4: There are otolaryngology departments in Teknon, Dexeus and Hospital Clinic. Which one would you prefer?  
 U4: Teknon.  
 S5: Dr. Ear is visiting at Teknon Tuesdays from 4 to 6 in the afternoon. Is it OK?  
 U5: Yes.  
 S6: Would you like me to find an itinerary there?  
 U6: Yes please.  
 S7: Would you prefer going by car or by public transportation?  
 U7: By public transportation, from Pl. Catalunya.  
 S8: Take FGC L7 from Pl.Catalunya and get out at Sarria.  
 U8: Thank you.  
 S9: Thank you for calling.

In insurance web pages, specialists are usually associated with the formal names of the specialization, which can present a difficulty for people without medical knowledge. In order to solve this problem the CA would allow the user to provide a common name for doctor specialization, for example *nose doctor*, *heart doctor*. To implement this functionality, we have parsed the Physician Specialist Glossary on the WebMD site ([8]) and build a knowledge source in the XML format.

An example of dialogue in this scenario supported by our system is given in Dialogue 1.

In order to improve the collaborative ability of the CA, the information on the specialists presented to the user would include information on how to go to the specific address or equipment where the she/he works. In particular, the system could guide the user about how to reach an address in three different ways: by public transportation (itinerary is retrieved by accessing Barcelona city web page on public transportation,[9]), by car and walking (using the information provided by Google Maps).

## 2) Searching for practical drug information

Practical information about a particular drug is usually available, either on the prescription information sheet or online. However, this information is not easy to read for all patients. The paper sheet that comes inside the drug box can be lost or printed in small fonts that can result difficult to read for elderly people. There are already online resources giving information on medical prescription of most drugs, but computer skills are needed to access them. A CA could be useful for answering most common questions included in drugs prescriptions, and the linguistic and conceptual resources required by the CA for this task are limited.

Prescriptions in many countries follow the same official structure. This structure favors answering most common questions related to drug prescriptions. In many countries there are online databases prescriptions of the drugs allowed in the country and those prescriptions follow the official structure. Those databases are usually accessible in the country language (for example, the Spanish database related to Spanish drugs, [10], accessible in Spanish language).

The sections in the drugs prescriptions in Spain are the following:

1. What is DrugName?
2. Before using DrugName
3. How should you take DrugName?
  - 3.1 What happens if you miss a dose? .
  - 3.2 What happens if you overdose?
4. What I should avoid while taking DrugName?
5. Conservation of the DrugName

The typical drug prescription is structured in sections that are associated with most common questions. For this reason, most users questions can be just matched to those in the

prescription and information answering it can be easily extracted. For example, a typical question the user can ask the system would be: *“I missed a dose of tardyferon. What should I do now?”*. Since there is a subsection associated with the question *“What happens if you miss a dose?”*, the system will just have to extract and present the fragment of text associated with this question.

### 3) Searching for an Open Pharmacy in the Neighbourhood

Finding an open pharmacy in the neighborhood is a quite common situation. Barcelona city hall's web page provides lot of relevant information about public health care ([11]), including all city pharmacies (together with their address, business hours and telephone numbers). In order to help the user in such scenario our system would just ask him the time and the place, parse the information about the pharmacies in the web page and present specific details of the pharmacies close to the user's location, indicating those that are open at that time. The system would use information about the addresses in the same zone (also available in the city hall's web page).

### 4) Stating a Date to Visit your Doctor

We have also considered a CA could be useful not only to help the user to find web information but also to perform other tasks, such as that of using a transactional web service to help the user when stating a date to see a doctor.



Figure 2. Making an appointment to see your doctor

In this scenario, the system will assist the user to give the data the service needs. The main challenge in this scenario, as in many other using transactional services, is that mistakes could have negative consequences. For this reason, clarification and confirmation dialogues are needed.

In particular, we have adapted our system to assist the user when using the web site of the public health care provider in Catalonia ("CatSalut"), that allows the user to make a reservation for an appointment to a primary health care centre (CAP) by introducing the user's personal health care card number, as shown on Figure 2. Once the user has introduced this number, then a calendar showing possible dates is shown. Finally, when a correct date has been selected a drop down menu displaying time for the selected day is shown.

The CA would facilitate the reservation of a date and time by allowing the user to introduce NL sentences expressing time, what can result friendlier for several users. For example, the user can introduce *“Please program for me a visit to my doctor on Monday or Wednesday, but I am free only after 18:00.”* and the system will parse the sentence and will automatically select the day and the time within the user specified choices.

## B. The Prototype

In order to study the main challenges the described scenarios present, we have developed a prototype using limited linguistic and dialogue resources. The prototype only supports text-based communication in Spanish. The prototype consists of three modules: the input/output module, the Dialogue Manager and the Task Manager. The system also includes a conceptual knowledge base (CK) that represents the application domain knowledge involved in communication. The CK is shared among all the system modules. These modules are described next

The input/output module processes user interventions and presents system messages. In the current implementation, this module consists of a Chabot that uses pattern matching templates for parsing the input request. Those patterns are stored in a XML file. The input is matched against those patterns and, if the match is found, the right command is forwarded to the Dialogue Manager.

The Dialogue Manager determines the content of the system intervention. It controls the information the Task Manager needs from the user to access the web and also how the information obtained from the web has to be presented to the user. In current implementation, this module is quite simple.

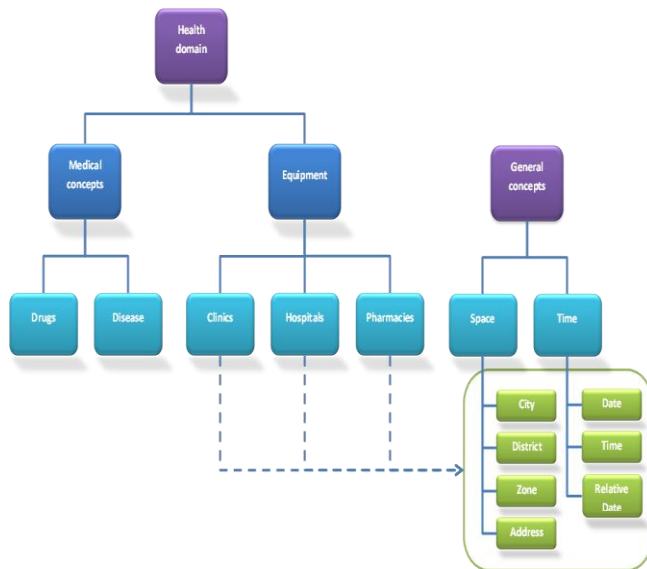


Figure 3. The Domain Conceptual Knowledge

The Task Manager controls the web access. Once the content of the user intervention has been passed to the Task Manager, it chooses which web site to automatically interact with. For accessing the web sources automatically, we have chosen HtmlUnit ([12]) because it is an open source Java library that creates HTTP calls which imitate the browser functionality. To extract the information from these web sources we developed a template based algorithm. This algorithm uses most common tags in simple html pages (such as those defining tables and lists) to extract the information.

One of the main challenges of our work consists of the representing the information obtained from the several web sources involved in a conversation with the user in an efficient form. For this reason, the CK, where information from several web sources is integrated plays a central role in our system. The CK involved in the specific studied scenarios (described above) has been represented in an ontology, shown on Figure 3.

As mentioned in the introduction, the use of ontologies can also be used to improve the collaborative abilities of our system when presenting the results of a web search in two different situations: when there were no results and when there were too many results. These two situations are critical in several of the scenarios we have considered, in particular, when searching a new specialist and when finding a specific equipment.

When assisting the user to search for specific information, in case the CA finds no results, concepts in the query can be substituted by their upper classes (to relax query constraints). For example, the system can replace a specific address in a query by the more general concept zone. In case too many objects satisfying the user's goal are found, the resulting set can be classified considering domain knowledge, as done in Dialogue 1 when the system propose the user to search for specialist in big hospitals (equipments are classified considering their size).

### III. CONCLUSION AND FUTURE WORK

In this paper, we have described an intelligent CA to guide the user when accessing several types of web sources in the medical domain. We have studied several scenarios where the user needs access to web medical resources of different type: glossaries, informational and transactional services. We have also considered scenarios where the combination of several web resources is needed.

We have implemented a prototype of the conversational assistant in form of a Chabot guiding the user in Spanish in the scenarios described: searching information (an specialist, an open pharmacy, a drug prescription), as well as performing a transaction (stating a date to see the doctor in a primary health care centre). In several of these scenarios, the system could, additionally, guide the user to go to a specific address (or equipment) presenting three alternative ways: using public transportation, a car or walking. The prototype incorporates an algorithm we implemented that extracts the information needed from the set of predefined web pages.

Future work will include extending the system to deal with new scenarios, even complex scenarios implying the integration of information obtained from several web sites using automatic web service composition solutions. We are also planning to extent the linguistic and dialogue system resources to support more complex communication phenomena in several languages. Additionally, we want to study how to adapt the system to different types of users.

### REFERENCES

- [1] I. Okoye, J. Mahmud, T. Lau, and J. Cerruti, "Find this for me: Mobile information retrieval on the open web," in International Conference on Intelligent User Interfaces (2011), 2011, pp. 3–12.
- [2] J. Allen, G. Ferguson, N. Blaylock, D. Byron, N. Chambers, M. Dzikovska, L. Galescu, and M. Swift, "Chester: Towards a personal medical advisor," *J.Biomedical Informatics*, pp. 500–513, 2006.
- [3] D. Sonntag, R. Engel, G. Herzog, A. Pfalzgraf, N. Pflieger, M. Romanelli, and N. Reithinger, "Handheld Multimodal Interaction with Ontological Knowledge Bases and Semantic Web Services". Springer, Heidelberg, 2007, pp. 272–295.
- [4] D. Guzzoni, C. Baur, and A. Cheyer, "Active : A unified platform for building intelligent web interaction assistants," in Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops), pp. 417-420.
- [5] Talk project.  
[http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ\\_RCN=6527011](http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=6527011), <retrieved: December, 2011>.
- [6] M. Gatus and M. Gonzalez, "A flexible dialogue system for enhancing web usability," in Proceedings of the 18<sup>th</sup> international conference on World Wide Web. Madrid: ACM, 2009, pp. 1167–1168.
- [7] M. Gatus and M. Gonzalez. "Using Domain Knowledge for Fostering the Collaborative Ability of a Web Dialogue System," *NWeSP'2011*.Salamanca, 2011.
- [8] [http://doctor.webmd.com/physician\\_finder](http://doctor.webmd.com/physician_finder), <retrieved: December, 2011>.
- [9] <http://w110.bcn.cat/portal/site/Mobilitat>, <retrieved: December, 2011>.
- [10] <http://www.prospectos.net>, <retrieved: December, 2011>.
- [11] <http://w3.bcn.cat/sanitat/>, <retrieved: December, 2011>.
- [12] <http://htmlunit.sourceforge.net/>, <retrieved: december, 2011>.

## Face Detection CUDA Accelerating

Jaromír Krpec

Department of Computer Science  
VŠB – Technical University Ostrava  
Ostrava, Czech Republic  
krpec.jaromir@seznam.cz

Martin Němec

Department of Computer Science  
VŠB – Technical University Ostrava  
Ostrava, Czech Republic  
martin.nemec@vsb.cz

**Abstract**— Face detection is very useful and important for many different disciplines. Even for our future work, where the face detection will be used, we wanted to determine, whether it is advantageous to use the technology CUDA for detection faces. First, we implemented the Viola and Jones algorithm in the basic one-thread CPU version. Then the basic application is widened to the multi-thread CPU version. Finally, the face detection algorithm is also implemented for the GPU using CUDA technology. At the end, final programs are compared and the results are presented in this paper. For our future plans, the speed-up of face detection is very important. By supporting CUDA technology, the process of the face detection showed considerable speed-up.

**Keywords**-CUDA; GPU; Face Detection; Viola and Jones algorithm

### I. INTRODUCTION

Face detection in images is quite complicated and a time-consuming problem, which found use in different disciplines, e.g., security, robotics, or advertising. By computer performance, disciplines of image processing, such as face detection, have significantly improved and progressed.

Even on current hardware, face detection is very time consuming, especially at the moment when large images are used. It is the same problem when we recognize faces in real time, for example from a camcorder. This is why the detection process must be accelerated.

In the last few years, graphic cards are increasing in performance; actually, the graphics processing unit (GPU) has greater performance than a classic central processing unit (CPU). Today, a graphic card can be used not only for rendering 2D or 3D graphics, but it can also be used for varied, especially parallel computations, which are not connected with the original task of graphic cards-rendering.

Compute Unified Device Architecture (CUDA) [1] technology is used to speed up the process of face detection, therefore we moved the main computation to the graphic card. Then, the final implementation was compared with a similar one-thread CPU and multi-thread CPU programs.

The main targets of this work are

- To implement the Viola and Jones algorithm [2] for the multi-thread CPU application,

- To implement the Viola and Jones algorithm for the GPU,
- To compare the speed of detection of individual programs, especially depending on the input image size, and
- To summarize and discuss results.

### II. RELATED WORK

There is much work which describes methods of face detection. There are methods which are based on template matching, skin detection and other techniques.

The algorithm of skin detection looks for areas covered by the skin color. Then, these areas could be marked as a face after fulfillment of other conditions (shape). Y. Chen and Y. Lin [3] widened this method. They added hair detection. If area of the face and area of hair are connected, the face is detected.

Template matching is a method for finding small parts of an image which match a template image containing face. Skin detection and template matching also can be combined into one method [4].

Next group of methods is based on the machine learning algorithms. The Viola Jones algorithm belongs to this group and it was chosen as the algorithm for acceleration in this work. The main reason why this face detection algorithm was chosen is the system of how this algorithm works. Thanks to using detection windows and Haar features, it offers a few of ways, of how to parallelize the detection process. The next reason is that there are many algorithms for face detection based on Viola and Jones.

A few works are also written about acceleration object classification with some good results. For example, in the work by C. Gao and S.L. Lu reached for image size 256 x 192 pixels 37 frames/sec for 1 classifier and 98 frames/sec for 16 classifiers [5].

### III. VIOLA AND JONES ALGORITHM

The Viola and Jones algorithm was used for the face detection, which was divided into two parts. The first part is for training a set of classifiers. For this process, MIT CBCL Face Database [6] was used. This database contains 2429 images with face and 4549 no-face images. Based on these

samples weak classifiers are created. The second part it is the detection itself, when faces are detected in the input image.

A. Features

Haar features are used for computing feature values during training and detection. A weak classifier is always described by a Haar feature that was chosen during the training process.

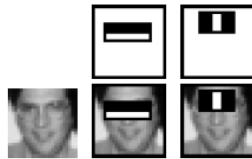


Figure 1. Using Haar features in the input image [2]

Every classifier consists of a black part (marked as B) and a white part (marked as W), and these parts cover some area of the image. Pixels are joined with one of these two parts of features and the final feature value for the current area of the image is calculated as:

$$f(x) = \sum_{w \in W} x(w) - \sum_{b \in B} x(b) \tag{1}$$

That means:

$$II(x, y) = Data(x, y) + II(x - 1, y) + II(x, y - 1) - II(x - 1, y - 1) \tag{2}$$

Data represents the original image, II is the integral image and (x,y) is the current position in the image.

B. Integral image

The integral image is an (m+1) x (n+1) multidimensional array created from an input image with m x n dimensions, where every value is counted as a sum of pixels in the interval (0...I - 1) x (0...I - 1). So:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \tag{3}$$

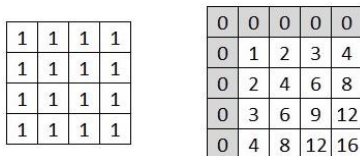


Figure 2. On the left, original image data; on the right, integral image

Now, the sum of pixels for a selected area can be easily quantified with constant speed:

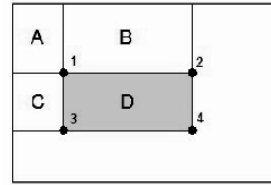


Figure 3. Sum of pixels in integral image

The value in point 1 is made of the sum of pixels in area A, the value in point 2 is A+B. For point 3, the value can be determined as the sum of areas A + C and in point 4 it is equal to the sum of all four areas A+B+C+D. By knowing the area of D, the corner point values could be used to determine the sum of pixels in this area: 4+1-(2+3).

Integral image computation is continual scanning of the full input image. Nevertheless, this operation could be parallelized, especially in case that large data is prepared for the input.

The principle of counting is in divided into two parts. Because results of pixel summation in rows do not influence themselves (and it is same for columns), it is possible to count pixel sums in rows and pixel sums in columns from the new data. During computing more rows or columns will be processed then.

At first, parallel processing of all rows is completed:

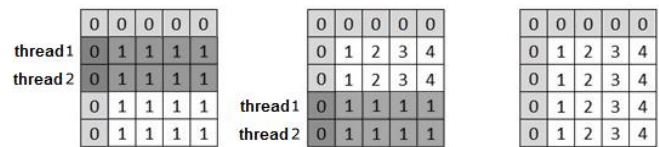


Figure 4. Parallel computing of rows

Then, the final integral image array is gained by the parallel computing of columns:

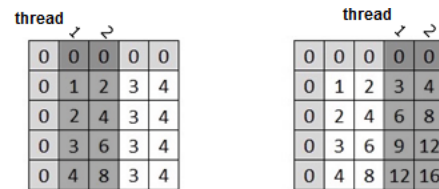


Figure 5. Parallel computing of columns

C. Detection

The detection process has a few steps. At first, the file with trained classifiers is loaded and the input image is also it is loaded, where the application will detect faces. This image is transformed to grayscale and creates the integral image and square integral image. These images are used for computing a standard deviation.



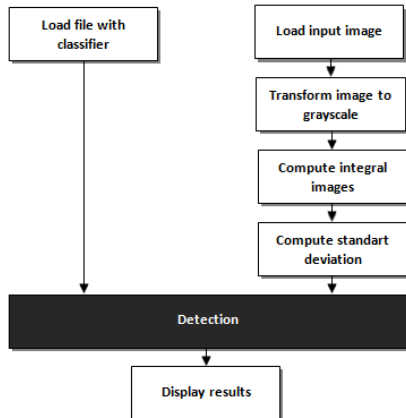


Figure 6. Detection steps

It then runs the main process of detection. During this operation, the parts of the image-detection window are counted. The minimal size of the detection window is the size of the images, which were used during the training.

The detection window is moved through the whole image and tries to determine if the current window could be a face. At the moment, when the window is in the last position in the image, the size of this detection window is enlarged by some rescale coefficient. Then, again, the full image is processed by the transformed detection window from the start position.



Figure 7. Detection window process

The situation is dependent on the feature value of the detection window, if the window is marked as a face or not. A standard deviation must be computer for every window, which is used for the feature value. Thanks to the integral image, we can count this operation with a constant time and it does not depend on the size of the window.

From the idea of the detection window, we can say that the computation time is affected by the count of trained classifiers, but especially by the image size.

It is obvious that the image size will increase the time calculation, because it must test more detection windows. Thanks to a greater image size, there are also more frequent transmissions of information between the device and host application.

The next table shows how the count of detection windows depends on the input image size (144 x 192 pixels) with the scale coefficient of 1.2.

TABLE I. COUNT OF DETECTION WINDOWS DEPEND ON SCALE

Scale	Detection windows
1,2	20449
1,44	4897
1,72	4480
2,07	4081
2,48	1617
2,98	748
3,58	589
4,29	286
5,16	128
6,19	44
<b>Result</b>	<b>37319</b>

D. Cascade algorithm

The cascade algorithm created by Viola and Jones brought some improvements in detection speed. Decreasing the time which is need for detection is based on the condition that there are more areas that do not contain faces. That is why it is not necessary to test all classifiers.

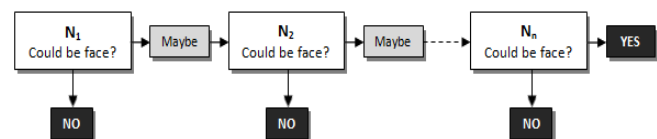


Figure 8. Cascade algorithm

The same MIT CBCL Face Database was used for the training cascade classifiers. The final count of strong classifiers is 15.

Each one of these strong classifiers contains a group of weak classifiers. The total sum of weak classifier is 529. It is significantly more than using one strong classifier. But for most detection windows, the algorithm expects that all weak classifiers will not be tested.

TABLE II. COUNT OF WEAK CLASSIFIERS FOR EVERY STRONG CLASSIFIER IN THE CASCADE

Stage	Number of classifiers	Stage	Number of classifiers
1	2	9	35
2	9	10	40
3	16	11	52
4	21	12	54
5	23	13	57
6	27	14	62
7	27	15	71
8	33	<b>Total</b>	<b>529</b>

Table 2. shows the sums of weak classifiers for every strong classifier in the cascade.

IV. PARALELL PROCESS

Basically, there are three possibilities how to parallel detection process:

A. Detection windows

Because the final results of feature values do not depend on the other detection window final results, it is easy to parallelize the detection windows. In the same moment we can test more detection windows. This could be realized on a CPU with more threads, but we can also use the GPU.



Figure 9. Parallel windows detection process

The problem occurs at the moment, when the detection window is in the last position of the current scale and it is necessary to rescale it. With the rescaling, we also create a different set of features, which must also be rescaled. The problem is when the other threads do not finish their detection and they are still in the old scale. At the moment, when features are rescaling, other threads will use a bad set of features and it could make bad detections.

One of the solutions could be that every thread will have its own copy of the features. However, from the memory view, this solution is not good enough. That is why we implemented a different solution. It uses only one set of features and it is shared by all the threads. To prevent problems, these threads are synchronized in the moment, when they got through the full image. Then the detection window and features are rescaled and threads are executed again with a changed set of features.

This is the method from these three that was chosen for implementation and testing.

B. Weak Classifier

The second way is in the parallel processing of Haar features. It signifies that only one detection window in the same moment is tested by more features.

If only one strong classifier is used, then the application runs parallel through a full set of weak classifiers. In case, when it used set of strong classifiers, only weak classifiers in one stage will be parallel processed. After a positive result of the current strong classifier is achieved, a new set of weak classifiers is loaded.

C. Scale detection window

If detection windows are in the last image position, windows change size for the next process. But there is no dependence between the scaling of different detection windows.

For example, if a smaller detection window detects a face in some position that does not mean that a larger window will detect the face in the same position. So, it is possible to have in one set of detection windows in one thread. This is

not a very good solution for very large images, because for the small sizes, it could create many windows, therefore, making memory demanding for common hardware.

V. IMPLEMENTATION

CUDA is technology developed by the NVidia Company, which can be used for diverse demanding computations on the GPU.

CUDA uses kernels that are executed n times in every thread and the thread is identified by the specific number.

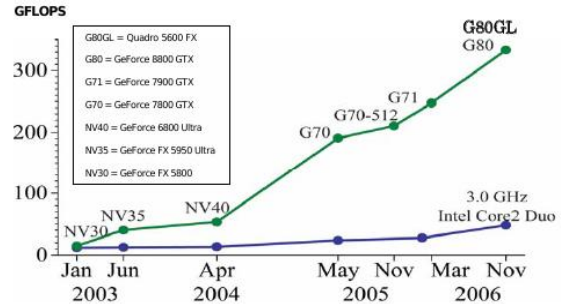


Figure 10. Compare Intel CPU with nVidia GPU [7]

The architecture of CUDA consists of grids, which are divided into smaller units - blocks. The hardware has a group of multiprocessors and it assigns each block to a multiprocessor. And finally, blocks consist of threads. These smallest units can be synchronize in one block.

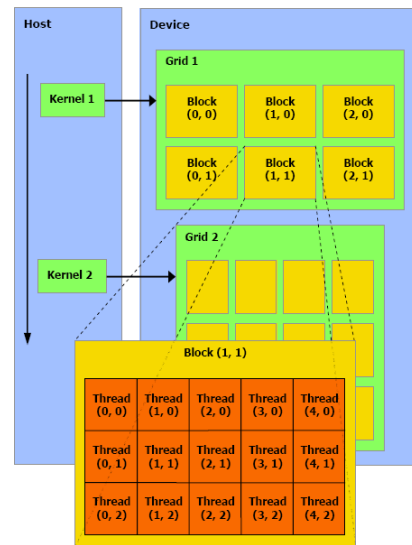


Figure 11. CUDA architecture [7]

In general, the CUDA program starts with memory allocation in the device, while data on the host are prepared. Then, the data are copied from the host to the device. Because the copying from host to device and from the

device to host is a time-consuming process, it is necessary to limit sending data.

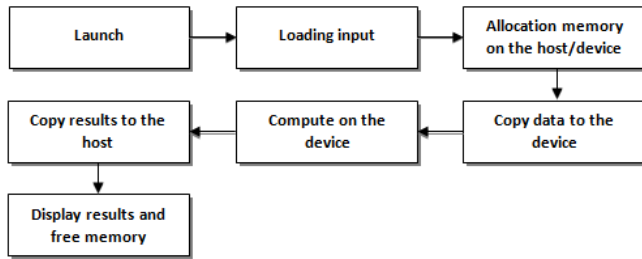


Figure 12. Typical CUDA program

After the data are ready on the device, it is possible to launch kernels. When the computation is finished, results are copied back to the host.

Finally results can be displayed and the allocated memory is released.

The scheme of the program is the same as Figure 12 shows.

The GPU implementation is similar to CPU multi-thread application. The idea is the same. I copy only what is needed to the device - integral images, trained classifiers and detection windows.

For every detection window's scale, a CUDA kernel is executed. The first version of the program computes positions of detection windows in the client application. A set of windows with the same size is computed. Then it is sent to the device and the detection process for the current windows can start. Of course, the detection window has same size, but the position depends on the index of thread. After that last detection window in the kernel is tested, the results are sent back to the host and the information about new detection windows is prepared on the host. This process is repeated until the scale reaches the final value.

The transmission between client and device is time-consuming, what shows a small adjustment.

For the next version, on the client side a count of detection windows and size is computed. This information is sent to the device. Now, it is possible to compute the position of a current detection window based on the received data from the client and index of thread. This adjustment caused an acceleration of detection an average of 15 times.

In the final implementation, a GPU computation of integral images is also used, which is also described in this paper.

## VI. RESULTS

At first, during testing I progressively compare a program with one-strong classifier with a multi-strong classifiers program. For both these implementations, a one-thread and multi-thread variation was created. Then these programs are compared with a GPU program.

### A. Integral image

Because the integral image is also computed by parallel threads, the following graph presents the results. It shows how the time needed for the computation depends on the image size.

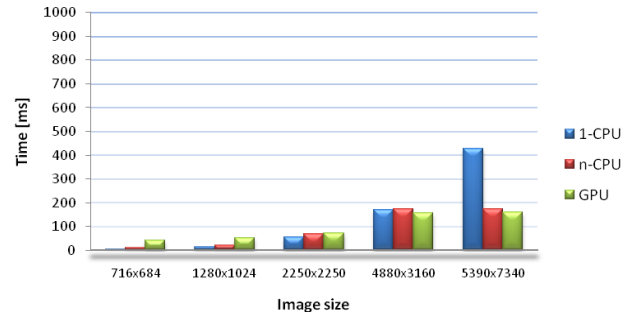


Figure 13. Integral image comparing

From the results we can see, that time computation is lowest for the GPU implementation, while the CPU program is significantly slower.

Nonetheless, the integral image computation is not as time-demanding as detection.

### B. One strong classifier detection

We then tested a one-strong classifier face detector. The following graph shows how long the process of detection takes for different image sizes. For one-strong classifier, 200 weak classifiers were trained from more than 5000 input samples.

It is not a cascade variation and that is why it is necessary for every detection window to test every weak classifier from the trained set.

For testing, three different image sizes were chose: 716x684, 1280x1024 and 2250x2250 pixels.

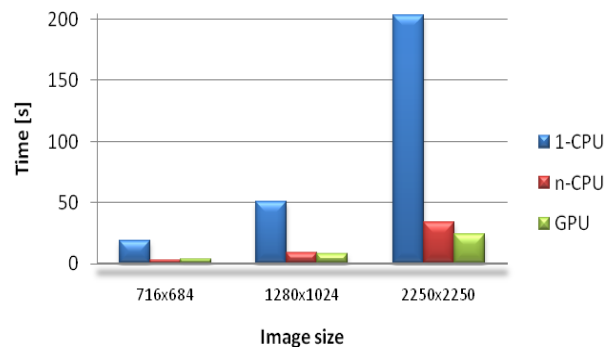


Figure 14. One-strong classifier detection

The results of one-strong classifier detection are 18.44 s for one-Thread for the 716x684 image size, 51.048 s for 1280x1024 pixels and 203.752 s for 2250x2250 pixels.

Eight threads were used for multi-threading testing and the results are 3.128 s, 8.694 s and 33.791 s. For the GPU program it is only 3.495 s, 7.808 s and 24.255 s.

### C. Cascade detection

The final comparison is with regards to the cascade variant of the Viola and Jones algorithm. The input file with classifiers contains 15 strong classifiers. Each one of these has a group of weak classifiers.

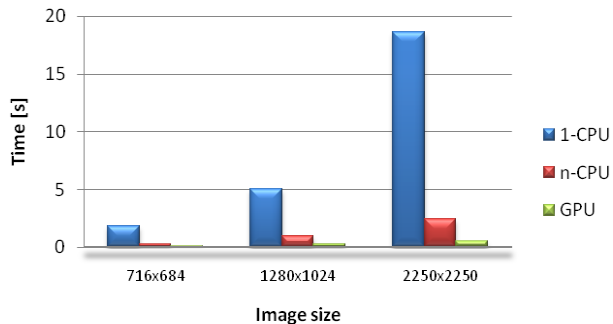


Figure 15. Cascade detection

The result shows that if it is not need test all classifiers set for current detection window, the computation is certainly faster than one strong classifier variation.

In the last graph we can see that one CPU thread is again the slowest from all the implementation modes, but this was obvious from the algorithm principle. For one CPU thread program it takes only 1.827 s, 5.008 s and 18.644 s. The next testing is with the multi-threading algorithm and results are 0.263 s, 0.989 s and 2.423 s. Finally, the GPU implementation takes only 0.117 s, 0.256 ms and 0.530 s. The final result is the same like the program with one strong classifier, so the GPU detection is the quickest from the presented forms of implementation.

## VII. CONCLUSION AND FUTURE WORKS

The possibility of multi-thread and implementation of the Viola and Jones face detection algorithm were presented here. It is not only about CPU thread applications, but especially about GPU.

All programs were tested and it shows that thanks to using threads the face recognition process can be accelerated against the basic one CPU version.

From the graphs, we can see that the detection time depends on the image size; this is the main factor. For the computation of the integral image, GPU implementation is the fastest.

In the next tests, the result is that the one thread CPU variant is obviously slower than the multi-thread CPU and GPU implementation.

From the test results, it is convincing that the GPU detection is usable with reasonable time-consuming results against the CPU variants. It is possible to see that the GPU detection is an average of 35 times faster than one thread CPU detection. In comparison to the multi-thread CPU variant, the results are closer, but the GPU is still quicker.

For the future, it is planned to widen the face detection capabilities for the possibility of recognizing faces with faces stored in a database. The input images will be gained from a camcorder. The real-time detection is the reason why the speed of face detection is very important. The output from the detection will be sent for the comparison with saved faces.

## REFERENCES

- [1] <http://developer.nvidia.com/category/zone/cuda-zone> <retrieved: January, 2012>
- [2] M. Jones and P. Viola, Robust Real-time Object Detection, International Journal of Computer Vision, vol. 57, no. 2, 2004, pp. 137-154,
- [3] Y. Chen and Y. Lin, Simple Face-detection Algorithm Based on Minimum Facial Features, 33<sup>rd</sup> Annual Conference of the IEEE Industrial Electronic Society(IECON), Taipei, Taiwan, November 2007, pp 455-460
- [4] S. Tripathi, V. Sharma, and S. Sharma, Face Detection using Combined Skin Color Detector and Template Matching Method, International Journal of Computer Applications, vol. 26, no. 7, 2011, pp. 5-8,
- [5] C. Gao and S. L. Lu, Novel FPGA based haar classifier face detection algorithm acceleration, In Proceedings of International Conference on Field Programmable Logic and Applications, 2008, pp 373-378
- [6] <http://cbcl.mit.edu/cbcl/software-datasets/FaceData.html> <retrieved: January, 2012>
- [7] <http://www.behardware.com/art/imprimer/659/> <retrieved: January, 2012>

## User Attention in Mobile Devices

Pekka Isomursu

School of Music, Dance and Media  
 Oulu University of Applied Sciences  
 Oulu, Finland  
 e-mail: pekka.isomursu@oamk.fi

Minna Isomursu and Mari Ervasti

Technical Research Centre of Finland  
 VTT  
 FIN-90571 Oulu, Finland  
 e-mail: minna.isomursu@vtt.fi  
 mari.ervasti@vtt.fi

**Abstract**— The multichannel information (over)flow emerging in modern mobile multimedia devices creates a situation where the user’s attention is a valuable asset for which different channels have to compete. This paper explores issues that help in effective use and maintenance of the user’s attention level in such environment. We demonstrate with examples how the user’s attention level can maintain high at different stages of task switching and execution.

**Keywords**-mobile Internet; attentive user interface; user interface; UI; ambient notification; human factors; design.

### I. INTRODUCTION

Today’s mobile communication has come a long way from simple voice calls. People are blending channels and media in a device-agnostic way, using mobiles not only for communication but also for being informed and entertained as well as for creating and maintaining their social networks. Web2.0 is turning mobile. Users’ communication threads may include both synchronous and asynchronous communication that may not have clear temporal start and end points. The number of participants may vary during a single communication thread.

Digital interruptions and related disruptions have been extensively studied in the desktop world. In the desktop environment, the interruptions disrupt digital activities, as the user is working with PC and is rarely doing other activities. However, interruptions in mobile and ubiquitous environments interrupt us in our everyday tasks such as sleeping, working or enjoying the opera. Therefore, design of interruptions for mobile and ubiquitous communication technology requires special research and design attention.

On a related work from the desktop context, Jackson et al. [8] have studied the effect that the incoming flow of emails has in people’s work efficiency. They found out that the way most users handle incoming email causes far more interruption than what is commonly assumed. The common reaction to the arrival of an email message is not to delay the response to a time that is more convenient to the user but to react within six seconds. The time it took to recover from email interrupts and return to the interrupted work task at the original work rate was on average 64 seconds.

All this multichannel information (over)flow creates a situation where the user’s attention is a valuable asset for which different channels have to compete. Some see attention as new currency for business and individuals [3].

In our current research, we are looking at novel ways to minimize disruption in mobile devices. This paper explores issues that help in effective use and maintenance of the user’s attention level in a mobile environment. We first briefly look at theoretical models of attention in Section II. Then, in Section III we discuss the principles of minimizing disruption and maximizing comprehension that help in the design of alerts and notifications, and the tools that we have identified to support those principles. In Section IV, we describe use scenarios that exploit the principles and tools described in Section III. Finally, we draw conclusions and discuss possible future work in Section V.

### II. THEORETICAL BACKGROUND

Attention is the cognitive process of selectively concentrating on one aspect of the environment while ignoring other things. Human attention is very complicated both psychologically and physically. Current understanding of attention divides it into several phases and concurrent processes that co-operate in complicated ways. A clinical model by Sohlberg and Mateer [13] divides attention into five classes with growing difficulty of execution:

- **Focused attention:** responding discretely to specific visual, auditory or tactile stimuli.
- **Sustained attention:** maintaining a consistent behavioral response during continuous and repetitive activity.
- **Selective attention:** maintaining a behavioral or cognitive set in the face of distracting or competing stimuli.
- **Alternating attention:** shifting the focus of attention and moving between tasks with different cognitive requirements.
- **Divided attention:** responding simultaneously to multiple tasks or multiple task demands.

An investigative model by Posner and Fan [10] includes the following attentive networks that coexist in the human brain:

- **Alerting network:** perceiving an incoming stimulus, which leads to a state of raised alertness.
- **Orienting network:** selecting relevant information from the stream of incoming stimuli, i.e., focusing attention to the sensory events.
- **Executive network:** exercising voluntary control over thoughts, feelings and actions, in order to maintain behavioral goals, i.e., maintaining attention.

Posner and Fan's model corresponds to the typical steps taken when the user switches tasks in the mobile context: 1) the user is alerted to pay attention to an event, 2) she orients her attention towards understanding what it is about, and then 3) decides how to respond and discards or executes the related task. As a simple example, the user pays alert attention when she hears the beeping of the phone and flashing of an email icon on the device screen. At a suitable moment, she then orients her attention to check the information content of the alert, e.g., sender and the topic of the email. Executive attention comes to action when she starts reading and processing the content. Each of these three attentive networks has its own role in the attention process and they can act simultaneously.

### III. MINIMIZING DISRUPTION, MAXIMIZING COMPREHENSION

Alert and notification design needs to balance between minimizing the caused disruption and at the same time, maximizing comprehension. A notification causes an interruption, which may cause disruption to whatever the user is doing. The level of disruption depends on several things, e.g., user context, how focused the user is on her current task, how strong the notification signal is, how similar the notification is to the current task, and what interaction modality (sound, visual, vibration, etc.) is used for the notification.

The user interface (UI) and the notification itself should support the user in responding to the notification in an appropriate manner. Response options should be clearly presented, and the desired action should be easy to perform. Similarly to alerts and notifications, also the execution of user responses can be designed to disrupt as little as possible, giving the user the possibility to continue with the primary task as smoothly as possible.

Alerts and notifications cause interruption (prompting transition and reallocation of attention focus) that can cause disruption. For the user, the disruption can mean the following:

- It draws the user's attention away from current activity.
- It requires cognitive processing and/or physical activity.
- It makes it difficult to return to the primary task.

Additionally, in mobile environments the mobile device can interrupt and disrupt not only the user of the device, but also other people that are physically or virtually nearby. This can create socially challenging (e.g., embarrassing) situations and tension.

Notification cues that take into account the needs of both the user as well as other people in the vicinity of the user are both subtle and public [5]. Successful subtle and public cues can help avoiding attention overload and misinterpretations in social context. Often they enable a combination of different kinds of notification cues, including new ways for devices to notify their users that also enable people to express themselves in various ways.

The following tools can be identified in minimizing disruption and maximizing comprehension of notifications and alerts:

- Filtering
- Choosing modality
- Interruptibility status of others
- Attentive UI, and
- Support for task switching.

In the following paragraphs, we will discuss these tools in detail.

**Filtering** can be done by setting priorities and rules on the handling of notifications. Simple examples are the rules to filter spam out of incoming emails and directing certain emails to different folders [11]. One might also control the rhythm of the information flow, e.g., by having clear intervals between non-urgent notifications so that the user is not disrupted continuously, and thus is able to concentrate on her primary task more efficiently. The eWatch system [12] is a wrist watch with vibration motor and sensors. This wearable device is aware of user activities and context. It combines an interruptibility measure, email priority level, and cybersensor data to decide how and when to notify the user.

**Choosing modality** of the notification cue according to a user's context or foreground task. The device could identify user's context, for example, according to measured ambient noise levels or usage level based on time since the user's last interaction with the device, and choose the most optimal modality to present notification cues. Also, research suggests that the notification creates most disruption if it is very similar to the modality of the foreground task [4], e.g., when the user is in the middle of the conversation, speech cues would be more disruptive than, for example, visual color coded cues.

Brown and Kaaresoja [2] have investigated vibrotactile messages that communicate multidimensional information in mobile phone alerts. They enable communication of more information through phone alerts, information to be communicated discretely without disturbing others, and alerting users in noisy environments.

Ambient soundscape by Jung and Schwartz [9] subtly notifies mobile users in a multi-user environment by offering an artificial sound environment and ambient notification service. For example, an incoming email is announced by enhancing background music with a personal non-speech audio cue (favorite notification instrument) that is played from the loudspeakers near the user's current position with increased volume. The service avoids distraction of other people, increases privacy and confidentiality, and prevents

the audio cue to be identified as a notification cue by outsiders.

**Interruptibility status of others.** Users of mobile phones tend not to be aware of the status of interruptibility of the person they are trying to contact. There is also limited freedom in choosing alternative channels of interruption, and channels that do exist do not allow for any subtlety of expression. The system could set the appropriate notification level, e.g., message, vibration, private knock or public ring, of the user’s cell phone and also show the user’s level of haste [14]. This profile would then be available to the users’ contact network, thus helping callers to choose an appropriate level of interruption.

**Attentive UI.** With Attentive UI, the device exploits the measurement of user’s attention level and focus. It can, e.g., alert the user for a non-urgent notification only when it realizes the user is not concentrating on another task and after that show more information on the location where the user’s attention is focused. The concept of attentive UI is utilized in AuraOrb [1]. It is an ambient notification display that uses social awareness cues, such as eye contact to detect user interest in an initially ambient light notification. Once detected, it displays a text message with a notification heading visible from any direction. Touching the orb causes the associated message to be displayed on the user’s computer screen. When user interest is lost, AuraOrb automatically reverts back to its idle state.

**Support for the switching** between the old and the new task, i.e., for alternating attention. Recovery tools can help the user to re-focus attention easier and faster. These tools can include, for example, visual indicators of occluded application windows, saving and displaying task context at the time of disruption, and playback of actions [7].

We next describe use scenarios that exploit the principles and tools described above.

#### IV. MOBILE SCENARIOS

The first scenario described below, “Gaining User Attention”, supports subtle alerting. The second scenario, “Supporting task switching”, helps the user orient on the new task and rapidly switch to the Execution Attention mode.

##### A. Scenario: Gaining User Attention

Angela is meeting her friends, with her monoblock phone in her pocket when a message arrives (Fig. 1). The phone recognizes that she has the phone in a tight pocket and thus vibrates gently, in a sequence that indicates it’s a non-urgent message and not, e.g., a phone call. If she had baggy trousers the skin contact would be weaker and the vibration would be stronger.

At her convenience, Angela picks up the phone. The device recognizes that it is been held in hand, and checks if there is a face in the view of its camera, eyes facing the phone. If this is the case the device automatically displays the message details.

In a variation of this scenario, Peter is in a meeting, with his foldable phone opened up on the table. A message arrives. The phone beeps quietly and changes its screen color. Peter recognizes that a message arrived but is in the

middle of a lively discussion. After a few minutes there is a suitable micro break and Peter looks at the phone. With its secondary camera facing Peter, the device recognizes the gaze and displays message details.

Additionally, if the device has enough processing power and good-enough camera to identify that the gazer is Peter, this would add security to the system as the message would not be shown to others that might glance at the device display.

##### B. Supporting task switching

Jill is in the middle of editing a Smiley image when her business acquaintance John calls. On display (Fig. 2), Jill sees pointers to John’s business information, her correspondence with John as well as the document she wanted to discuss with her next time they speak. Just by glancing at all this, she gets better oriented back to her business with John. If she desires so, before answering she can also use a precious moment to open and check some related data. This can be done fast since everything is readily available behind the shown links.

After the call, Jill switches back to her image editing. But now she is disoriented as her mind still wanders on what she discussed with John. To help her orient back into her previous task the device shows on the screen an animation of the last edits she made (Fig. 3). Now her thoughts are back to her previous task

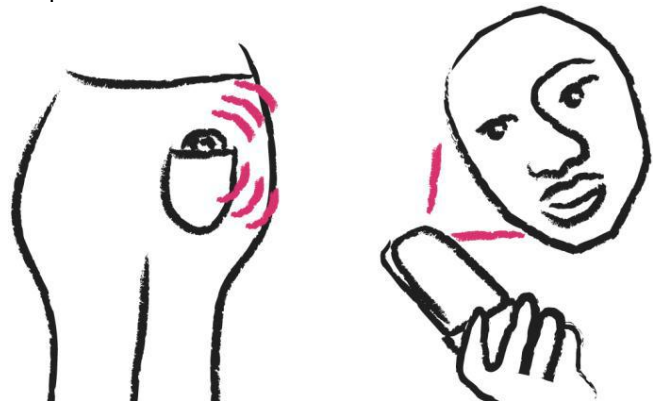


Figure 1. Gaining and recognizing user attention.

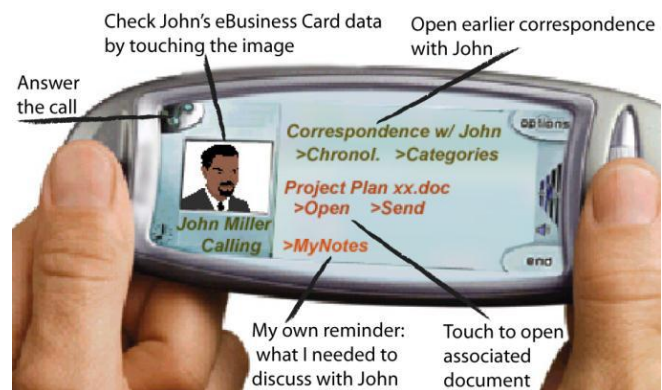


Figure 2. Associative UI.



Figure 3. Animating last edits of user's disrupted task.

## V. CONCLUSION AND FUTURE WORK

To tackle the mobile multichannel information overflow, we have shown examples about how user attention can be supported at different stages of task switching and execution. In our future work we wish to extend the tools described in Section III and their usage in the mobile environment.

We would also like to better understand the pitfalls in the user experience of these technologies. For example, on one hand, due to its appeal to the basic instinct of survival humans have a natural inclination to seek for moving, bright, eye-catching, and rapid access to information even if it distracted them. Even when we know what is best for us we do not always act accordingly, as is demonstrated, e.g., by Jackson's observation [8] that most people let the constant flow of email distract their work, or by the common habits of smoking and unhealthy diet. On the other hand, being focused requires effort, willpower and practice, and we easily lose our attention to spontaneous thought. Designers of attentive UI's may face a problem of balancing between calm and attentive UI's that allow us to focus, and our natural inclination for craving constant stream of new, mobile and flashy interruptions that create the illusion of being connected and informed, but at the same time corrupt our possibilities to concentrate and focus.

In this paper, we have focused on switching from one task to another. In the future, we would also like to study divided attention, i.e., how to support situations where people wish to simultaneously execute multiple tasks.

## ACKNOWLEDGMENT

We gratefully acknowledge the support to our work from the following sources: Nokia Corporation and, at Oulu University of Applied Sciences, both Focus Area of Internet and ICT (Mr. Kimmo Paaajanen) and the InnoMajakka project (Ms Sanna Oja), funded by the European Social Fund of the European Union.

## REFERENCES

- [1] Alto Saar, M., Vertegaal, R., Sohn, C., and Cheng, D. (2006) AuraOrb: Social Notification Appliance. Proc. ACM SIGCHI Conference on Human Factors in Computing Systems CHI 2006, April 22-27, Montréal, Québec, Canada, pp. 381-386.
- [2] Brown, L.M. and Kaaresoja, T. (2006) Feel Who's Talking: Using Tactons for Mobile Phone Alerts. Proc. ACM SIGCHI Conference on Human Factors in Computing Systems CHI 2006, Montréal, Québec, Canada, pp. 604-609.
- [3] Davenport, T. and Beck, J. (2001) *The Attention Economy: Understanding the New Currency of Business*, Harvard Business School Press.
- [4] Gillie, T. and Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research* (50), pp. 243-250.
- [5] Hansson, R. (2001) Subtle and Public Notification Cues for Mobile Devices. Proc. Ubiquitous Computing: International Conference, Ubicomp 2001, Atlanta, Ga, USA. Springer. Pp. 240-246.
- [6] Hudson, S.E. and Smith, I. (1996) Electronic Mail Previews Using Non-Speech Audio. Proc. ACM SIGCHI Conference on Human Factors in Computing Systems CHI 1996, Vancouver, Canada., pp. 237-238.
- [7] Iqbal, S. and Horvitz, E. (2007) Disruption and Recovery of Computing Tasks: Field Study, Analysis and Directions. Proc. ACM SIGCHI Conference on Human Factors in Computing Systems CHI 2007, San Jose, California, USA, pp. 677 - 686.
- [8] Jackson, T. W., Dawson, R., and Wilson, D. Understanding email interaction increases organizational productivity. August 2003/Vol. 46, No. 8 *Communications of the ACM*. pp. 80-84
- [9] Jung, R. and Schwartz, T. (2007) A Location-Adaptive Human-Centered Audio Email Notification Service for Multi-user Environments. *Human-Computer Interaction, Part III*, Proc. 12th International Conference on Human-Computer Interaction HCI 2007, Beijing, China, pp. 340-348.
- [10] Posner, M. I. and Fan, J. (2001) Attention as organ system. J. Pomerantz (ed), *Neurobiology of Perception and Communication: From Synapse to Society the IVth De Lange Conference*. Cambridge, UK: Cambridge University Press.
- [11] Röcker, C., Bayon, V., Memisoglu, M., and Streitz, N. (2005) Context-Dependent Email Notification Using Ambient Displays and Mobile Devices. *Active Media Technology 2005 (AMT 2005)*, pp. 137-138.
- [12] Smailagic, A., Sieworek, D., Maurer, U., Rowe, A., and Tang, K. (2005) eWatch: Context Sensitive System Design Case Study. Proc. IEEE Computer Society Annual Symposium on VLSI, ISVLSI 2005, pp. 98-103.
- [13] Sohlberg, M. M. and Mateer, C. A. (2001) *Cognitive Rehabilitation: An integrative neuropsychological approach*. New York: Guilford Press.
- [14] Vertegaal, R., Dickie, C., Sohn, C. and Flickner, M. (2002) Designing Attentive Cell Phones Using Wearable EyeContact Sensors. Proc. ACM SIGCHI Conference on Human Factors in Computing Systems CHI 2002, Minneapolis, Minnesota, USA, pp. 646-647.



## Semiautomatic Evaluation of Websites Usability

Eva García, Antonio García, Luis de-Marcos, Salvador Otón, José-Ramón Hilera

Computer Science Department

University of Alcalá

Alcalá de Henares, Spain

eva.garcial@uah.es, a.garciac@uah.es, luis.demarcos@uah.es, salvador.oton@uah.es, jose.hilera@uah.es

**Abstract** - This paper presents a semiautomatic evaluation method of usability in websites. Expert review is done in this case using a template based on the ISO 9241-151 guidelines that provides numerical and graphical results. Such results permit to observe the aspects of usability that are missed on the evaluated website. This template has many applications as it can be used both to evaluate websites and to teach heuristic evaluation method for students who are studying subjects of usability.

**Keywords** - Usability evaluation; usability guidelines; heuristic evaluation; web usability

### I. INTRODUCTION

Usability is one of the most important features to consider when making systems that have large audiences, which need to operate in an intuitive system, without prior training or direct support [1]. In addition, usability is the second most important problem (just after security) for the acceptance of systems by users [2]; so, it must not be omitted when developing computer systems. However, currently there are few tools to help automating the process of usability evaluation.

The aim of this paper is to show a new way of semiautomatic heuristic evaluation [3] of websites. For this, we have developed a spreadsheet template with guidelines to evaluate, which are organized into groups, and each of them is given a value based on a certain scale. When the scores of all guidelines are filled, a result will be obtained showing the more problematic aspects in relation to the usability of the evaluated website.

In this case, the heuristics are based on guidelines provided by ISO 9241-151 [4]. This standard is specifically a part of the ISO 9241 family of standards, which began publication in 1997 under the title "Ergonomics of human-system interaction" and provides requirements and recommendations related to the attributes of the hardware, software and environment that contribute to usability and ergonomics underlying principles.

Part 151 in particular, is titled "Guidance on World Wide Web user interfaces" and provides guidance for user-centered design of web interfaces in order to increase usability. Web user interfaces are aimed to all Internet users and to closed user groups such as members of an

organization, customers or suppliers of a company or other specific user communities.

User interfaces of different types of user agents such as web browsers or additional tools such as web authoring tools are not directly aimed in this part of ISO 9241, although some of the guidelines could be applied to these systems as well.

Web user interfaces can be presented in a personal computer, in a mobile system or in any other device connected to the network. While the recommendations contained in this part of ISO 9241 are applied to a wide range of technologies available in front-end, web interfaces design of mobile devices or smart devices may require additional guidance.

A new way of semiautomatic heuristic evaluation of websites is proposed here. To do this, Section II describes the design of the template, i.e., groups that are designed to bring together the guidelines, the scale that has been used to evaluate each guideline, the content of the template and how can results be interpreted. In Section III, template is applied to several websites and obtained results are presented in each case, as well as a comparison of them. Finally, some conclusions and future research are presented.

### II. TEMPLATE DESIGN

To evaluate in a semiautomatic way the usability of a website, a spreadsheet template has been created from an existing one created by the company Userfocus, available at [5]. The original template consists of 247 guidelines created by the company, grouped in nine categories: home page usability, task orientation, navigation and IA, forms and data entry, trust and credibility, writing and content quality, page layout and visual design, search usability, and help, feedback and error tolerance.

The new template (Figure 1) proposes a significant improvement over the previous one, including new guidelines from ISO 9241-151 standard. These guidelines have been put together in new groups, also different from the original template, and also taking into account the sections of ISO 9241-151 standard.

In addition, the rating scale for each guideline has also been modified to take more realistic the scoring system.

A. *Designed groups*

As mentioned earlier, the guidelines of the template are based on ISO 9241-151. This standard consists of the following sections:

1. Scope
2. Normative references
3. Terms and definitions
4. Application
5. A reference model for human-centered design of World Wide Web user interfaces
6. High-level design decisions and design strategy
7. Content design
8. Navigation and search
9. Content presentation
10. General design aspects

Navigation		
Guidelines		
8.2.1	Making navigation selfdescriptive	2
8.2.2	Showing users where they are	-1
8.2.3	Supporting different navigation behaviours	1
8.2.4	Offering alternative access paths	1
8.2.5	Minimizing navigation effort	0
8.3.2	Choosing suitable navigation structures	1
8.3.3	Breadth versus depth of the navigation structure	1
8.3.4	Organizing the navigation in a meaningful manner	-2

Figure 1. Overview of the template showing a category with scored guidelines.

Only Sections 7 to 10 have been taken into account for the template, because they are the only evaluable guidelines for a website already developed. This is because Sections 1 to 5 contain general information about the ISO and Section 6 contains high level decisions and design strategies, so they will only be applicable to websites that are still in development but not for those that have already been developed.

Guidelines for evaluating usability have been divided into several groups according to the different sections of the ISO 9241-151 listed above, but with an exception: Section 8 has been divided into two parts, as in the ISO the aspects of navigation and search are put together in the same section, and it is preferable to evaluate separately the navigation and the search of the website.

Therefore, sections that the template finally contains are as follows:

1. Content design: it is about everything related to the conceptual model of content, content objects and functionality.
2. Navigation: it deals with issues related to the structure and components of the navigation.
3. Search: it is about the aspects related to search terms on a web page.

4. Content presentation: it treats the aspects related to the design of a web page, links design, interaction objects and text design.

5. General design aspects: the guidelines of this group aim to be designed for a culturally diverse and multilingual use, and provide help and make web user interfaces error tolerant. In addition, it cares about the names of the URLs, the time of downloading, designing with independence of input devices, etc.

Each one of these groups will be a category in the spreadsheet template, and each category includes the guidelines for the subject matter that group.

B. *Rating scale*

Original template had a scale of integer values between -1 and +1, i.e., the possible values that could be given to each guideline were:

- 1: Guideline is not satisfied.
- 0: Guideline is not totally satisfied (it is partially satisfied).
- +1: Guideline is satisfied.

If guideline is not satisfied, in the original template the value should be left blank. This way is not very intuitive, and it does not allow intermediate values when a guideline is almost completely satisfied or it hardly satisfies.

To improve this system of evaluation of the guidelines, our template uses the Likert scale [6]. This is reasonable since guidelines consist of a title (which is what appears in the template and corresponds to a section of the ISO), but also an explanation (available within the ISO), which sets out statements that should be satisfied.

To do this, it is necessary that the scale has an odd number of levels. It should be noticed that the guideline may not be applicable. Due to this reason, a 5-level scale (with 5 categorical variables) has been chosen, i.e., the possible values that can be given to each guideline are:

- 2: Strongly dissatisfied.
- 1: Dissatisfied.
- 0: The guideline is not applicable.
- +1: Satisfied.
- +2: Strongly satisfied.

C. *Template content*

The template consists of 7 categories or sections, of which the last five correspond to designed groups that have been explained in Section 2.1. The categories that make up the template are:

- Instructions: it contains instructions for using the spreadsheet workbook and to properly fill the template. To do this, the values that can be assigned to each guideline are explained, i.e., the rating scale discussed in Section 2.2.
- Results: after filling in all the scores of the guidelines, it will be necessary to see this category for interpreting the obtained results. This category consists of a table summarizing the results for each set of guidelines, as well as a chart where results of the previous table can be graphically seen.

- Content design: it is the category corresponding with the design group called “Content design”.
- Navigation: it is the category corresponding with the design group called “Navigation”.
- Search: it is the category corresponding with the designed group called “Search”.
- Presentation: it is the category corresponding with the designed group called “Content presentation”.
- General design: it is the category corresponding with the designed group called “General design aspects”.

D. Interpretation of results

In the category “Results” of the template, results of the evaluation of the website can be interpreted. This will have a table that identifies, for each of the designed groups, the following information:

- Raw score: it is the addition of the scores of each of the guidelines of the group in question.
- #Questions: it is the number of guidelines that are evaluated in the group in question. Guidelines that are not rated are not taken into account (but there should not be any, because all values must be filled) nor those that contain a 0 (because in that case guideline is not applicable).
- Score: it is the score in percentage terms, i.e., it gives an idea of the degree of compliance with a given set of guidelines. Its value can vary between -100% (if none of the guidelines of designed group is satisfied) and +100% (if all guidelines of the designed group are completely satisfied). It is calculated as follows:

$$\frac{\sum_{i=1}^n v_i}{2 \cdot n} \cdot 100 \tag{1}$$

where n is the number of guidelines of the group in question, and vi is the rated value to each of the guidelines.

- Overall score: it is the overall score in terms of percentage, i.e., it gives an idea of the degree of general compliance with the guidelines and, therefore, of the ISO 9241-151. It is the average of the scores obtained in the score categories.

In addition to the results table, this section contains a radar chart that helps getting a quick idea of the general usability of the website evaluated:

- The larger the shaded area, best usability. In a general way, a website has a good usability when the score of each section is at least 50%. A 30-50% range means a reasonable usability and a score under the 30% means that the website has a poor usability.
- The higher the scale of the graph, best usability average score.
- The angles in the chart represent each of the designed groups as described in 2.1, so that the greater the score has an angle, better usability will have the corresponding group.

III. REAL APPLICATION

The tool is created for helping the heuristic evaluation, since several evaluators are involved in the heuristic evaluation but the template can only be filled by a single evaluator at a time. The idea is that each assessor completes the template on his own, and then pooling the results to issue a final report. This report is the only document the user will receive, because it is written in an easy to understand vocabulary for the user, so the evaluation sheets will not be shown to the user, because they contain technical vocabulary that he or she could not understand.

To demonstrate the operation of the template as a heuristic evaluation method, we have chosen two websites where it has been applied: a website that has been evaluated negatively and one that has been evaluated positively. To check the capabilities of the template, evaluations have been carried out by an usability expert who has experience in web interfaces (as recommended by Nielsen [7] to identify most of the problems).

A. Example of website evaluated positively

The site chosen for this section has been the London transport website [7]. After conducting the heuristic evaluation with the template, results obtained can be seen in Figure 2.

Summary of Results			
	Raw score	# Questions	Score
Content design	23	15	77%
Navigation	46	26	88%
Search	19	17	56%
Content presentation	42	37	57%
General design aspects	19	13	73%
Overall score		108	70%

Figure 2. Table of results of the heuristic evaluation of the London transport website.

As it can be seen in that figure, the main conclusions can be drawn are:

- The best evaluated group is the navigation group, with 88%, indicating that the website navigation is very good.
- The worst rated group is the search group, with 56%, indicating that there would be desirable to improve the website search system.
- The final average score is 70%, so that we can deduce that the website usability is quite good.

B. Example of website evaluated negatively

The site chosen in this case was that of an Atlanta restaurant guide (<http://www.restaurantguideatlanta.com/>). After conducting a heuristic evaluation with the template, we got the results that can be seen in Figure 3.

As shown in the figure below, the main conclusions obtained are:

- The best evaluated group is the general design aspects group, with 28%, indicating that these aspects can be even improved.

- The worst rated group is the content presentation group, with -41%, indicating that in this respect the website is quite poor.
- The final average score is -12%, so that we can deduce that the usability of the website can (and should) be improved significantly.

Summary of Results			
	Raw score	# Questions	Score
Content design	-3	12	-13%
Navigation	-16	22	-36%
Search	1	19	3%
Content presentation	-26	32	-41%
General design aspects	5	9	28%
Overall score		94	-12%

Figure 3. Table of results of the heuristic evaluation of the website of Atlanta restaurant guide.

### C. Comparison of the results of example

To compare the results of the example, it is best to observe the charts of both websites generated by the template (Figure 4) after performing the heuristic evaluation of them.

As shown in the figure below, there are considerable differences in charts:

- The shaded area is higher in the first chart than in the second one, indicating a better general usability of the website.
- The scale of the first chart varies between 0% and 100%, while the second chart scale varies between -50% and 50%. This indicates that in the first chart the average score of usability is greater than in the second chart.
- In the first chart, navigation is highlighting, followed by the content design and general design aspects, and it can be seen that the aspects that can be improved are the search and the content presentation.
- In the second chart highlights the low overall score, especially as regards the content presentation.

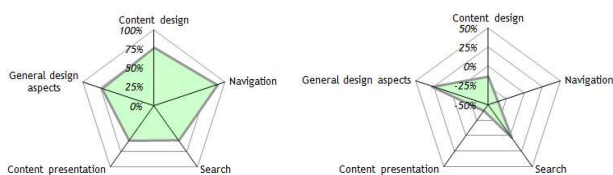


Figure 4. Graphic results of the usability evaluation of the London transport website (left) and Atlanta restaurants guide (right).

Comparing these results with those obtained in Sections III.A and III.B, same conclusions can be drawn when observing both the table and the chart obtained from the template, although it is easier and faster to get a general idea of the usability of a website by looking at the chart.

### IV. CONCLUSION AND FUTURE WORK

Once proposed and introduced the modifications in the original template, and having considered some websites to compare the results obtained, we can deduce that this is a

valid method for semiautomatic evaluation of websites usability. We will carry out a comparative experiment with this template and the original, so we will be able to deduce the real benefits of the improvement.

One of the advantages of using this template is that it is fulfilling a very important standard of web usability; and another advantage is that it is a semi-automated method, since the expert introduces values he deems appropriate and the template calculates the percentages of compliance of the guidelines and returns results with which to draw conclusions easily. However, it also has the disadvantages inherent to the heuristic evaluation, among which are the following: evaluators should be experts in usability and, if possible, be familiar with the type of interface that is being evaluated [8], and must have multiple evaluators [9] (which is expensive).

As future research we plan to adapt the guidelines of the ISO 9241-151 to mobile devices, i.e., research what guidelines can be used to heuristically evaluate the usability of websites viewed from mobile devices. Once this is done, the template will be modified and adapted to be used to evaluate the usability of mobile websites according to ISO 9241-151. We will also design a method to collect the results of the evaluation of all experts and achieve a common result.

Finally, we will try to automate the evaluation of some guidelines and analyze some user’s behavior patterns, like extracting the link structure of a website and obtaining the link graph according to a predefined ruleset, and then analyzing the behavior of the user while navigating.

### ACKNOWLEDGMENT

This research is funded by the FPI research staff education program of the “Junta de Comunidades de Castilla-La Mancha”. Authors also want to acknowledge support from the TIFyC research group and the Computer Science Department of University of Alcalá.

### REFERENCES

- [1] R. Gafni, “Quality Metrics for PDA-based M-Learning Information Systems”, in *Interdisciplinary Journal of E-Learning and Learning Objects*, vol. 5, 2009, pp. 359-378.
- [2] J. Buranatrived and P. Vickers, “An Investigation of the Impact of Mobile Phone and PDA Interfaces on the Usability of Mobile-Commerce Applications”, *Proc. IEEE 5th International Workshop on Networked Appliances*, 2002, pp. 90-95.
- [3] J. Nielsen, “Heuristic Evaluation”. <http://www.useit.com/papers/heuristic> (last access: 2011/09/06).
- [4] ISO 9241-151:2008. “Ergonomics of human-system interaction. Part 151: Guidance on World Wide Web user interfaces”. Organization for Standardization. Geneva, Switzerland, 2008.
- [5] Userfocus, “Expert Review Checkpoints”. <http://www.userfocus.co.uk/pdf/ExpertReviewCheckpoints.xls> (last access: 2011/09/06)
- [6] R. Likert, “A Technique for the Measurement of Attitudes”, in *Archives of Psychology*, vol. 140, pp. 5-53. New York, 1932 (cited by J.T. Cacioppo and G.G. Berntson, “Relationship Between Attitudes and Evaluative Space: A Critical Review, With Emphasis on the Separability of

- Positive and Negative Substrates”, in Psychological Bulletin, vol. 115, No. 3, pp. 401-423, 1994).
- [7] “Transport for London”, <http://www.tfl.gov.uk> (last access: 2011/09/06)
- [8] J. Nielsen, “Finding usability problems through heuristic evaluation”, Proc. SIGCHI conference on human factors in computing systems, pp. 373-380, Monterey, CA, USA, 1992.
- [9] J. Nielsen and R. Molich, “Heuristic evaluation of user interfaces”, Proc. SIGCHI conference on human factors in computing systems: Empowering people, pp. 249-256, Seattle, WA, USA, 1990.

# Motion-sound Interaction Using Sonification based on Motiongrams

Alexander Refsum Jensenius  
 University of Oslo, Department of Musicology  
*fourMs - Music, Mind, Motion, Machines*  
 Oslo, Norway  
 Email: a.r.jensenius@imv.uio.no

**Abstract**—The paper presents a method for sonification of human body motion based on motiongrams. Motiongrams show the spatiotemporal development of body motion by plotting average matrices of motion images over time. The resultant visual representation resembles spectrograms, and is treated as such by the new sonifier module for Jamoma for Max, which turns motiongrams into sound by reading a part of the matrix and passing it on to an oscillator bank. The method is surprisingly simple, and has proven to be useful for analytical applications and in interactive music systems.

**Keywords**-sonification; motion; motiongram; jamoma.

## I. INTRODUCTION

Motion and sound are closely linked in the real world, but not always so in interactive systems. Even though the awareness of sound has grown steadily since the early experiments on sonic interaction by, e.g., Gaver [1], [2], it is first in the last decade that the field of *sonic interaction design* has emerged as an established research field and design direction, as documented in, e.g., [3], [4].

A core challenge in sonic interaction design is to understand more about the relationships between action and sound, i.e., what types of sounds fit with what types of actions [5]. In the physical world, actions involving objects will always lead to some kind of sonic feedback dependent on the mechanical and acoustic properties of the actions and objects involved. Furthermore, there are countless examples of how motion and sound are part of a feedback cycle, where sound may again lead to action (e.g., dancing). In electronic devices, on the other hand, the sonic feedback (if there is one) is designed and constructed either mechanically or electroacoustically.

This paper will present one approach to understanding more about the interaction between motion and sound, and a method that can be used in the design process of interactive systems. The method is based on *sonification*, the representation of numerical data in an auditory form [6], of body motion captured using a regular video camera. Such an exploration of how it is possible to “translate” from motion to sound, or sound to motion, may give valuable insights into our multimodal cognition of both motion and sound, and may also be the starting point for explorations of systems using such relationships between motion and sound for various types of interaction.

The starting point for the paper was the observation that *motiongrams* (see Section III for an explanation) visually resemble spectrograms. I was therefore interested in exploring what would happen if motiongrams were turned into sound, as if they had been a spectrogram. The study has two aims:

- exploring how sound can be used in the analysis of music-related body motion
- exploring how sonification of body motion can be used in interactive systems

The paper starts with an overview of some related research. Then motiongrams are introduced, followed by an explanation of how motiongrams can be used to create sound. Finally, some examples of both analytical and interactive applications are presented and discussed.

## II. BACKGROUND

My approach to turning video images of body motion into sound is based on what could be called an “inverse spectrogram” technique. This was most directly inspired by the work on image *scanning* and *probing*, as proposed by Yeo and Berger [7], where an image is transferred into sound with frequency on the Y-axis and time on the X-axis. These techniques were later developed into *raster scanning* and the creation of *rastrograms* in [8].

The idea of translating an image into sound is not new. The perhaps earliest example of using a spectrogram-like approach to sonification was the *Pattern Playback* machine built in the late 1940s by speech researcher Franklin S. Cooper [9]. This system made it possible to “draw” shapes that could afterwards be played back as sound. The UPIC system by Iannis Xenakis, developed in 1977, made it possible to use a digital pen to draw on a computer screen [10]. This approach is nowadays available in the Metasynth software [11], and a simplified version in the demo patch *Additive Synthesis* shipping with the graphical programming environment Max/MSP/Jitter. An augmented reality version of the same idea was used in Golan Levin’s *Scrappler* [12], where objects put on a table are tracked using computer vision and used to control the sound synthesis.

Parallel to this development, and closer to my own approach, are the many attempts at creating systems for controlling sound through movement, e.g., in interactive art. An early example here is that of Erkki Kurenniemi’s

electronic music instrument Dimi-O (1971), using video input for controlling the sound synthesis [13]. Other notable examples include David Rokeby's *Very Nervous System* (1982-1991) and SoftVNS, both of which have been used in a number of interactive installations and dance performances. In the last decade, the availability of graphical programming environments like EyesWeb, Isadora, and Max/MSP/Jitter, have made it possible for artists to easily set up interactive systems based on video input. Many of these systems use motion detection to control either sound synthesis or samplers in realtime. One such example is Pelletier's direct mapping of motion flow fields to sound [14], and subsequent motion-sound mappings using Gestalt-based feature extraction [15].

Also related to my work, but starting from a different premise, is the sonification of clarinetists' performance actions [16], [17]. These projects use data from a marker-based infrared motion capture system as the point of departure for the sonification, something which makes it possible to select specific points on the body/instrument to sonify. As such, it is a more specialized technique than what I am proposing, but it still shows some of the potential in a successful sonification process of motion to sound.

### III. MOTIONGRAMS

An overview of creating a motiongram is shown in Figure 1. The process starts by reading a video stream and converting it into a greyscale image. In future research it would be interesting to also use the color information, but the current exploration has been done with greyscale images only. It may also be useful to do some simple image adjustments at this stage, e.g., changing the brightness and contrast, so that the video used for further analysis is as clear as possible (Figure 1.2).

The next step involves producing the *motion image* by calculating the absolute frame difference between subsequent video frames (Figure 1.3). Dependent on the quality of the original image, and the noise level in the image due to video compression, lighting, etc., it may be necessary to filter the motion image (Figure 1.4). This can be done through simple thresholding, or applying a noise removal algorithm to remove groups of few pixels. The motiongram is created by calculating the normalized mean value for each row in the motion image (Figure 1.5). This means that for each image matrix of size  $M \times N$ , a  $1 \times N$  matrix is calculated. Drawing these 1 pixel wide "stripes" next to each other over time results in a horizontal motiongram (Figure 1.6).

As opposed to a spectrogram in which the intensity in the plot is used to show the energy level of the frequency bands, a motiongram is simply a reduced display of a series of motion images. There is no analysis being done, the creation process is only based on a simple reduction algorithm. This has made the technique very useful in many different applications, as has been summarized in [5].

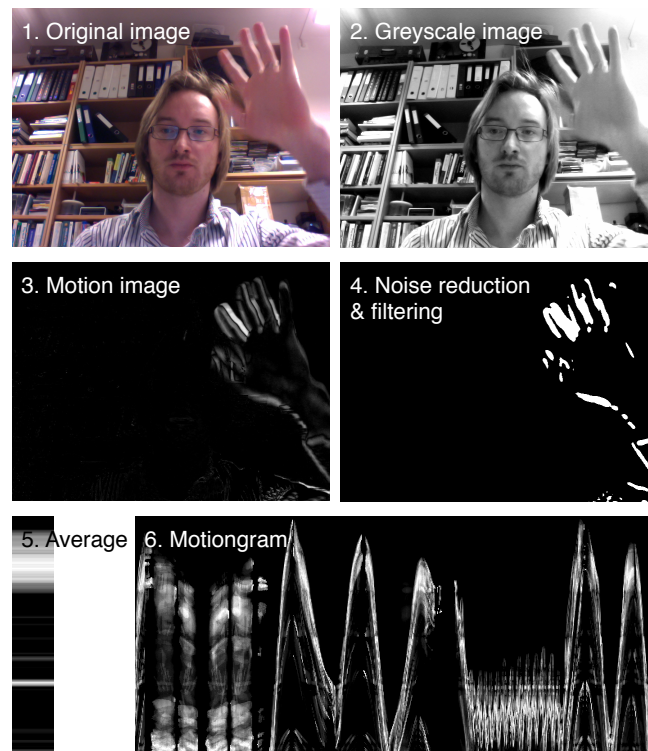


Figure 1. The steps involved in creating a motiongram: greyscale conversion (2), frame differencing (3), filtering (4), averaging (5) and plotting over time (6).

It is worth mentioning that a motiongram will only display motion in *one* dimension. Thus a horizontal motiongram visualizes only vertical motion, since all information about the spatial distribution of motion in the horizontal plane is represented by only 1 pixel for each row. When creating motiongrams it is therefore necessary to evaluate in which plane(s) the motion is occurring, before deciding whether to create a horizontal or a vertical motiongram (or both).

### IV. FROM MOTION TO SOUND

Since motiongrams share many visual properties with spectrograms, I was interested to see how they could be used as the basis for sonification of motion. The most obvious way of doing this is by treating the motiongram as a spectrogram, as suggested by Yeo and Berger in their scanning approach mentioned in Section II [7]. This way we can create a direct mapping from motiongram to spectrogram, as illustrated in Figure 2.

A minimal implementation of such an "inverse spectrogram" technique in the graphical programming environment Max/MSP/Jitter can be seen in Figure 3. The implementation is based on reading one line at a time from the motiongram matrix and turning this into an audio signal using the `jit.peek~` object. This is then sent to an interpolated oscillator bank (`ioscbank~`), which does the additive synthesis. The result is a direct sonification of the motion, where lower

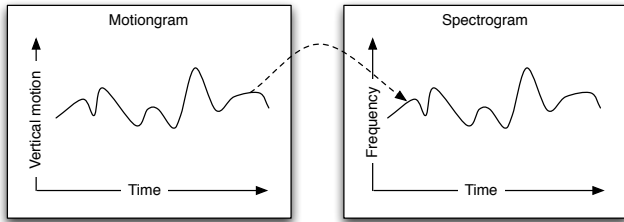


Figure 2. A direct mapping from motion data to spectral audio data.

sound frequencies are controlled by moving in the lower part of the image, and vice versa.

The sonification algorithm has been implemented in the module `jmod.sonifyer~` in the open framework Jamoma for Max [18]. Thus the module benefits from the extensive preset, mapping and cueing functionality present in Jamoma [19]. As for many other video modules in Jamoma, `jmod.sonifyer~` will adapt itself to any incoming matrix size, something which makes it easy to change between differently sized videos on the fly.

An example of how the sonifyer module may be used in conjunction with other Jamoma modules is shown in Figure 4, and a video tutorial of the functionality of the module can be seen in Video 1 (all video examples are available at [www.arj.no/sonifyer/](http://www.arj.no/sonifyer/)). The `jmod.input%` module gets video from the camera and passes it on to `jmod.motion%`, which calculates the motion image and does the noise reduction and filtering. The filtered motion image is passed on to `jmod.motiongram%`, which outputs a motiongram of the chosen size and direction (in this case horizontal). The right outlet of `jmod.motiongram%` passes the reference to the motiongram matrix on to `jmod.sonifyer~`, while the left outlet passes on the message of the internal counter in the motiongram algorithm. This counter keeps track of the column number that the motiongram is currently outputting, and is used to control the speed of the “playback” of the motiongram to sound. For realtime applications this counter increases for each new frame received from the camera (typically at 25 fps), and for non-realtime applications it can be used to scan through the image as fast as possible.

### V. EXAMPLES

Video 2 shows examples of sonification of some basic motion patterns: up-down, sideways, diagonal and circular. Here only a simple level of filtering and noise reduction is used, otherwise it is a direct translation from motiongram to sound. The examples show one of the largest problems with this approach to sonification: the motiongram’s ability to only display motion in one direction. Thus the up and downwards motion is clearly visualized in the motiongram, and heard in the sound, but all the other motion patterns (sideways, diagonal and circular) are not represented equally well with only one dimension being sonified.

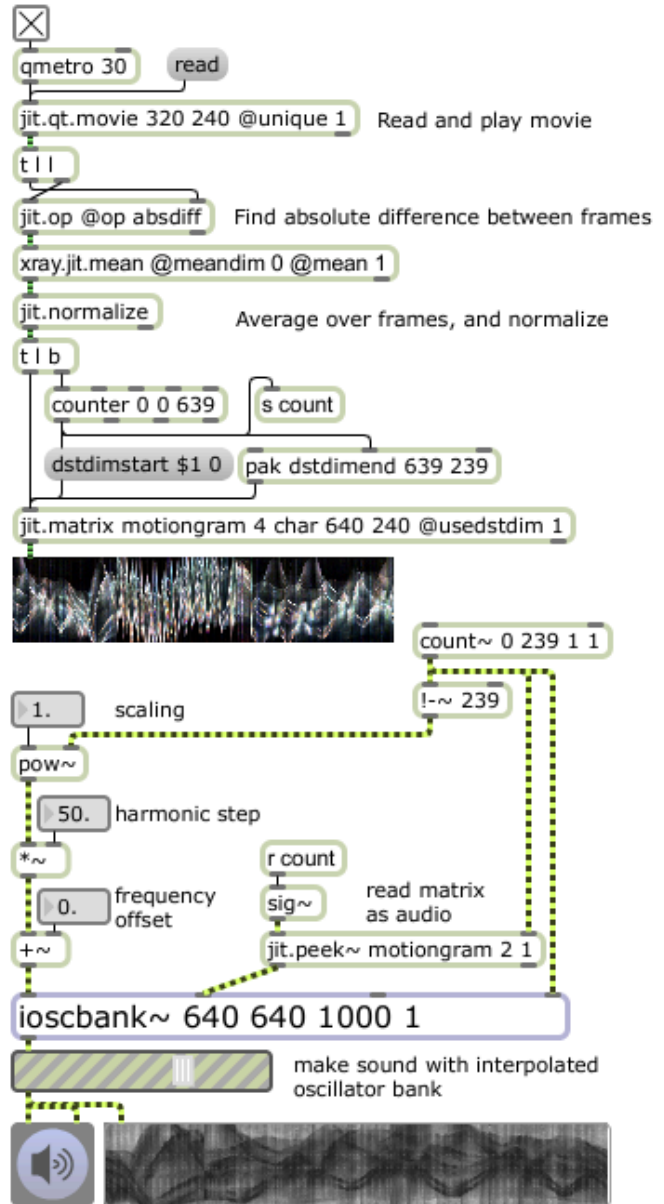


Figure 3. A minimal Max implementation for generating the sonification from video input. The oscillators are created in the `ioscbank~` object, and Jitter matrix data is converted to an audio signal with the `jit.peak~` object.

One attempt at sonifying the two axes at the same time is shown in Video 3. Here both horizontal and vertical motiongrams are created from the same video recording, and the sonifications of the two motiongrams have been mapped to the left and right audio channel respectively. While I originally thought this may be a good idea, the example shows that it does not work particularly well. Clearly more research is needed to find a better solution for sonifying the two dimensions.

Filtering and thresholding of the motion image is important for the final sounding result, as can be seen in



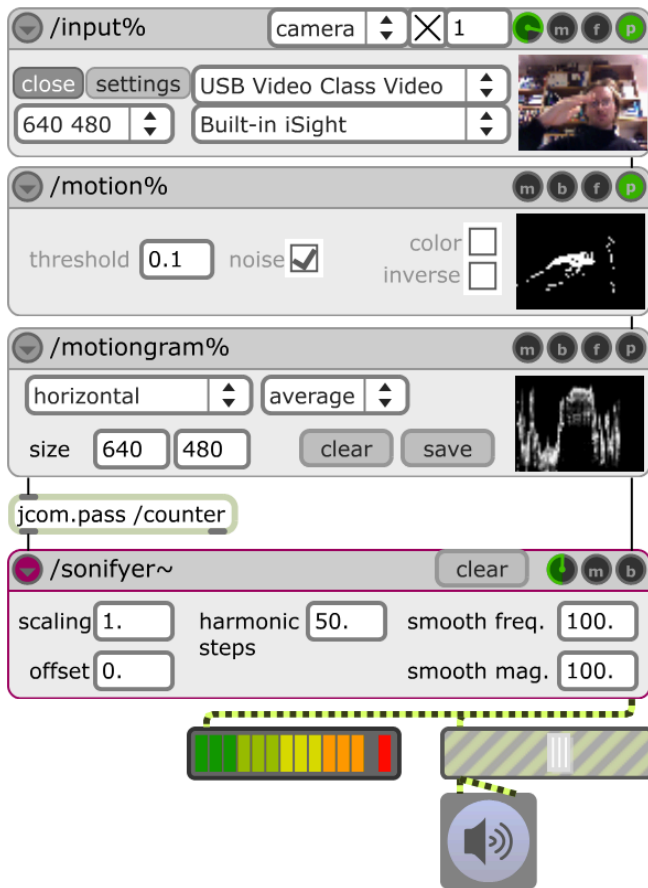


Figure 4. From the help patch of the `jmod.sonifyer~` module. The video input module is connected to a module creating the motion image and then to the motiongram module. Finally, the output motiongram is sent to the sonifyer module for the creation of sound.

an example of the sonification of a high-speed recording (200 fps) of a hand in motion in Video 4. Here three different types of filtering have been applied to show the different sonic results. When there is no thresholding and no noise reduction, all the details of the motion is shown in the motiongram and can also be heard in the sonification. Adding a binary threshold removes a substantial amount of pixels in the motiongram, and hence makes a cleaner sonification. Finally, adding a noise reduction algorithm further reduces the amount of pixels and sonifies only the most important part of the motion.

Video 5 shows an example of the sonification of a short violin improvisation. While the sonification manages to capture some of the details and temporal unfolding of motion over time, I generally find that a sonification of sound-producing actions tend to be confusing. This is probably because we expect that the sonified sound should be related to the sound-producing action. This, however, is not possible with such a generic sonification technique, which is based on translating all motion into sound without any prior

knowledge about the content of the video material.

A more successful sonification of the motion of a performer can be seen in Video 6 of a French-Canadian fiddler. Here we are focusing mainly on the clogging pattern that is created in the feet. The rhythmicity of this pattern is sonified clearly, and the change of rhythmic figure and tempo is easily audible halfway throughout the excerpt. See [20] for a more detailed analysis of this performance.

An example of the sonification of dance motion is shown in Video 7. First the original recording is shown, where a dancer moves spontaneously to a short musical excerpt, followed by a sonification of the same motion. Here the sonification of the motion shows some clear similarities to the sonic qualities of the original sound. This, however, is a special case of a good correspondence between the original sound and the sonification result. In general I would argue that sonifications should not be evaluated against the original sound, but rather against the motion that they are sonifying. It is only in cases of sound-imitating motion that the sonification will be similar to the original sound.

The sonification module has also been tested in music performance. An excerpt from a performance of the piece *Soniperforma* at Biermannsgården in Oslo on 18 December 2010 can be seen in Video 8, and a screenshot from the performance patch in Figure 5. This piece is based on applying only video effects to change the sonic quality. This way it is possible to create for example delays in the sound by applying a motion blur function on the video image.

## VI. DISCUSSION

The sonification technique based on motiongrams presented in this paper is still in development. While the method works well for some examples, there are also several issues that will have to be explored further:

*Dimensionality:* The limitation of handling motion in more than one dimension was shown in Video 2. This limitation is based on the fact that motiongrams average over each row (or column) and therefore reduce motion in the video from 2 to 1 dimension. I will continue to explore how it is possible to handle multiple dimensions (2 and 3) in sonification, but with an aim to continue keeping the process simple, direct and intuitive.

*Temporal resolution:* A challenge when working with video as the source material for a sonification process is the poor temporal resolution as compared to audio. This is particularly apparent when working with direct mappings from video to sound. For this reason it will be interesting to explore how high-speed (200-1000 fps) video recordings will work as the basis for sonification. Such frequencies are still far lower than the possibilities of audio synthesis, but may reveal some possible future uses of this type of sonification approach.

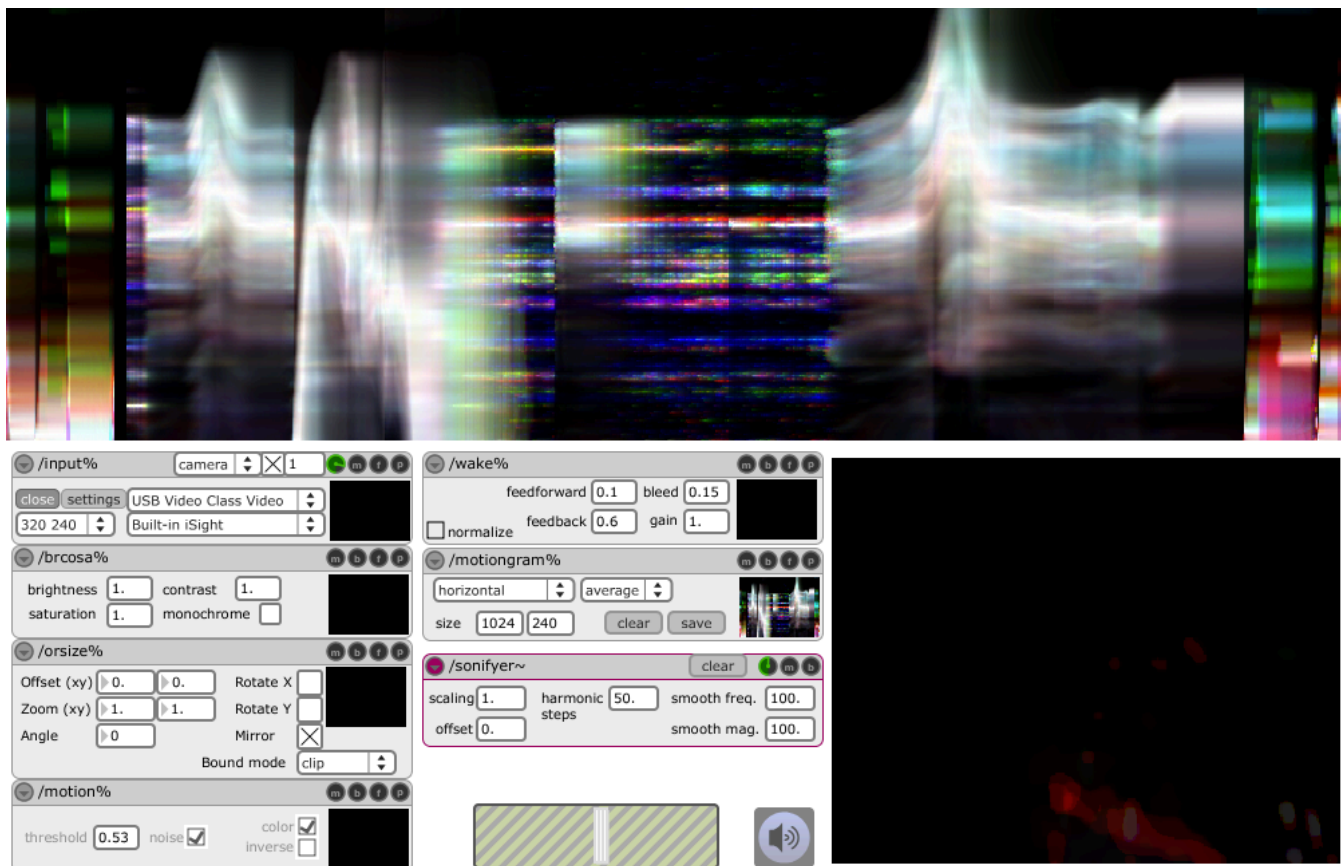


Figure 5. Performance patch using various types of video effects to modify the motiongram, and hence the output sound. The visual result from changing various video parameters can be seen in the motiongram at the top.

*Analysis:* The original idea of this sonification approach came from an analytic point of view: creating a tool to help in the analysis of various types of music-related motion. The exploration has shown that the largest potential of the method may be in the sonification of all sorts of non-sound-producing motion. If we were to create a sonification of the sound-producing action, it would be necessary to know where the excitatory parts of the instrument are in the image. This is a very different problem, and was never the intention of the project.

*Sonic interaction design:* I quickly realized that the current implementation probably has a larger potential in interactive than in analytic applications. The immediate and intuitive connection between motion and sound has opened for interesting sonic explorations in many different contexts, both in general human-computer interaction and for more creative applications. The method can be used as a tool to quickly create a sonification of body motion, which can later be used as the basis for designing a more complete sound design in an electronic device or system.

*Music applications:* As shown in Video 8, using video effects to modify sound in realtime has been a refreshing

approach to sound creation. It has been fun to perform with, and audience members have commented that the link between projected image and sonified sound works well. At first such a performance setup may seem odd, but in fact it is quite similar to performing with a regular instrument. Since sound is only created when there is motion, moving one hand in front of the camera can be used to excite the “instrument,” while the other hand can be used to modify the quality of the sound by changing video filters.

*Scalability:* The system has been tested on close-ups of hands, upper body, and full body video recordings. I have also done a test with a group of 20 students standing on the floor and being filmed from above. In such a setup it is possible to create a collaborative performance among the people making up “pixels” in the image.

*Stability:* The current implementation has been very reliable. The patch runs comfortably on a single laptop using a built-in camera, and can easily be extended to use any type of external camera. The video modules have been used in analytic and creative applications for the last 5 years, and have been adjusted so that they work well in all sorts of lighting conditions. Also, using the motion image as point of

departure means that the system does not rely on a particular type of background, as long as it is possible to make a separation between motion in the foreground and in the background.

## VII. FUTURE WORK

Issues to be addressed in future research include:

- optimizing the implementation so that it runs faster.
- implementing more interactive controls of the sonification parameters, e.g., based on extracted motion features.
- developing a non-realtime application. This would allow for creating more detailed sonifications of large motiongrams, e.g., of high-speed and high-resolution video material.
- exploring sonification of both horizontal and vertical motion, as well as from multiple cameras.
- exploration and user testing in many different contexts.
- exploring a similar approach to sonify data from infrared/inertial motion capture systems.

## ACKNOWLEDGMENT

This research has been funded by the Norwegian Research Council through the *Sensing Music-related Actions* project, and has been carried out in the fourMs lab at the University of Oslo.

## REFERENCES

- [1] W. W. Gaver, "Auditory icons: Using sound in computer interfaces," *Human-Computer Interaction*, vol. 2, pp. 167–177, 1986.
- [2] —, "The SonicFinder: An interface that uses auditory icons," *Human-Computer Interaction*, vol. 4, no. 1, pp. 67–94, 1989.
- [3] D. Rocchesso and F. Fontana, *The Sounding Object*. Firenze: Edizioni di Mondo Estremo, 2003.
- [4] D. Rocchesso, *Explorations in Sonic Interaction Design*. Logos, 2011.
- [5] A. R. Jensenius, "Action–sound: Developing methods and tools to study music-related body movement," Ph.D. dissertation, University of Oslo, 2007. [Online]. Available: <http://urn.nb.no/URN:NBN:no-18922>
- [6] S. Barrass and G. Kramer, "Using sonification," *Multimedia Systems*, vol. 7, no. 1, pp. 23–31, 1999.
- [7] W. S. Yeo and J. Berger, "Application of image sonification methods to music," in *Proceedings of the International Computer Music Conference*, Barcelona, 2005.
- [8] —, "Application of raster scanning method to image sonification, sound visualization, sound analysis and synthesis," in *Proceedings of the 9th International Conference on Digital Audio Effects*, Montreal, 2006.
- [9] F. Cooper, A. Liberman, and J. Borst, "The interconversion of audible and visible patterns as a basis for research in the perception of speech," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 37, no. 5, p. 318, 1951.
- [10] G. Marino, M.-H. Serra, and J.-M. Raczinski, "The upic system: Origins and innovations," *Perspectives of New Music*, vol. 31, no. 1, pp. 258–269, 1993.
- [11] "Metasynth [computer program]. available: <http://www.uisoftware.com/metasynth/index.php> (last checked: 08.11.2011)."
- [12] G. Levin, "The table is the score: An augmented-reality interface for real-time, tangible, spectrographic performance," in *Proceedings of the International Conference on Computer Music 2006 (ICMC'06)*, 2006.
- [13] M. Ojanen, J. Suominen, T. Kallio, and K. Lassfolk, "Design principles and user interfaces of erkki kurenne's electronic musical instruments of the 1960's and 1970's," in *Proceedings of the 7th international conference on New interfaces for musical expression*. ACM, 2007, pp. 88–93.
- [14] J.-M. Pelletier, "Sonified motion flow fields as a means of musical expression," in *Proceedings of the International Conference on New Interfaces For Musical Expression*, Genova, 2008, pp. 158–163.
- [15] —, "Perceptually motivated sonification of moving images," in *Proceedings of the International Computer Music Conference*, Montreal, 2009, pp. 207–210.
- [16] O. Quek, V. Verfaillie, and M. M. Wanderley, "Sonification of musician's ancillary gestures," in *Proceedings of the 12th International Conference on Auditory Display*, London, UK, 2006, pp. 194–197.
- [17] F. Grond, T. Hermann, V. Verfaillie, and M. Wanderley, "Methods for effective sonification of clarinetists' ancillary gestures," *Gesture in Embodied Communication and Human-Computer Interaction*, pp. 171–181, 2010.
- [18] T. Place and T. Lossius, "Jamoma: A modular standard for structuring patches in Max," in *Proceedings of the International Computer Music Conference*, New Orleans, LA, 2006, pp. 143–146.
- [19] T. Place, T. Lossius, A. R. Jensenius, and N. Peters, "Flexible control of composite parameters in max/msp," in *Proceedings of the 2008 International Computer Music Conference*, Belfast, 2008, pp. 233–236. [Online]. Available: <http://urn.nb.no/URN:NBN:no-20631>
- [20] E. Schoonderwaldt and A. R. Jensenius, "Effective and expressive movements in a french-canadian fiddler's performance," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, A. R. Jensenius, A. Tveit, R. I. Godøy, and D. Overholt, Eds., Oslo, Norway, 2011, pp. 256–259. [Online]. Available: <http://www.nime2011.org/proceedings/papers/G12-Schoonderwaldt.pdf>

# Touchscreen Interfaces for Visual Languages

Michael Hackett and Philip T. Cox  
*Faculty of Computer Science*  
*Dalhousie University*  
*Halifax, Nova Scotia, Canada*  
*e-mail: mhackett@cs.dal.ca, pcox@cs.dal.ca*

**Abstract**—Through the construction of a simple mock visual programming language (VPL) editor, this study compares two different styles of touchscreen interface and demonstrates the natural fit between touch input and visual programming. The touchscreen removes a layer of indirection introduced by the mouse and allows for a more direct relationship—literally “hands on”—with the objects on the screen. The addition of “multi-touch” also opens up intriguing possibilities for two-handed, immersive interfaces, with the potential for greater efficiencies than possible with the mouse’s single point of interaction.

**Keywords**—Touchscreens; visual programming; bimanual; user interfaces; kinaesthetic feedback

## I. INTRODUCTION

An often cited advantage of visual programming environments is the greater sense of *direct manipulation* that these environments provide, through the immediate feedback received while working with some graphical representation of the code. However, when the manipulation is performed with a mouse or a trackpad, as is typical in existing systems, an unnatural barrier is placed between the user and the software environment. Additionally, having to frequently switch hand positions and “operating modes” between mouse and keyboard slows users down and breaks their sense of flow, cutting into their productivity.

Recently, large touchscreen devices have become inexpensive and readily available, while new user interface designs are being developed that are tailored to touch input. Through these devices, users can interact more directly with their data, manipulating visual representations with the touch of a finger. And while the lack of a physical keyboard would seem to make these devices poorly suited for traditional text-based programming, the more direct control of on-screen objects that the touchscreen affords might just make them ideal for visual programming languages (VPLs).

This paper describes the design and implementation of a simple VPL code editor—codenamed “Flow”—for a touchscreen tablet device (an Apple iPad). The prototype was built in order to experiment with various user-interface options and to compare the usability and efficiency of the interfaces. The initial version has two different styles of interface, one based on the familiar drag-and-drop paradigm and the second using a novel bimanual (two-handed) approach. To

minimize the influence of language syntax on overall usability, the VPL is closely modeled on Prograph [1], an existing well known dataflow programming language. At this stage, the prototype is only a facade of a true visual programming environment—editing is very limited and execution is not yet supported.

## II. BACKGROUND AND RELATED WORK

Although there is little prior work to draw upon with respect to visual programming on a touchscreen, the design of the Flow prototype was informed by more general touch-input and user-interface research. Of particular interest and inspiration was research into bimanual interfaces, kinaesthetic feedback, and touchscreen gesture design.

### A. Two-Handed Input

Bimanual interfaces take advantage of the use of both of the user’s hands together, usually in an asymmetric fashion. Yee [2] and Wu et al. [3] suggest using the non-dominant (NP) hand to establish and maintain modes while the dominant (P) hand does work within that context (as suggested in Guiard’s Kinematic Chain model of bimanual action [4]).

This type of arrangement seems particularly well suited to today’s “multi-touch” displays (which can distinguish multiple simultaneous touches), and the larger screens make it feasible to create a touch interface that supports the use of two hands at once. Hence, it was decided early on to experiment with at least one interface of this style.

### B. Kinaesthetic Feedback

Sellen et al. [5] provide details on two experiments that compared kinaesthetic feedback (using a foot pedal) with visual feedback, and user-maintained feedback with system-maintained. They found that *actively* user-maintained kinaesthetic feedback significantly reduces mode errors (that is, instances in which a user performs an action appropriate for a mode that the application is not currently in), and allows for faster resumption of activity after an interruption. Wu et al. [3] use the term “kinaesthetically held modes” to describe this interface technique, while Raskin suggests the more succinct “quasimodes” [6, p. 55].

In contrast, “latching” modes, where a mode stays in effect until cancelled or changed, tend to lead to more mode errors [5][6], and users often struggle to figure out how to cancel the mode selection if they change their minds or realize they have made an error.

### C. Touchscreen Gesture Design

Nielsen et al. [7] have defined a number of principles and guidelines for designing gestures with usability and ergonomics in mind. They also presented a procedure for building a “gesture vocabulary” (the set of gestures in an interface) through user studies, then refining and testing the resulting gesture set. They stress the importance of keeping ergonomics in mind and warn against basing gesture design on the recognition capabilities of the hardware, as this may result in gestures that are illogical and “stressing or impossible to perform [for] some people”.

Mauney et al. [8] performed a user study with participants from eight different countries and found that there was strong agreement across cultures for gestures with a physical or metaphorical correspondence to the objects being manipulated (“direct manipulation gestures”), but fairly low agreement on gestures that were symbolic in nature. This suggests that symbolic gestures should be avoided (a point also made by Frisch et al. [9]), except perhaps as expert-level shortcuts.

## III. DESIGN PRINCIPLES

While far from an exhaustive list, this section describes three key principles that guided the design of the prototype.

*Discoverability:* Norman and Nielsen [10] have criticized today’s commercial touch tablet interfaces, citing numerous examples of gestures that are not easily discoverable or guessable, and are often learned only by reading about them elsewhere. To support discoverability in Flow, a toolbar displays buttons for all available commands, obviating the need for a series of hidden (and often arbitrary) gestures that need to be learned. On the other hand, moving graphical objects by dragging them with a finger is intuitive and easily discoverable; it is not necessary to create a special button for this action.

*Minimizing Mode Errors:* Flow uses kinaesthetic feedback, in the form of screen contact, to regulate modes. A mode becomes active when a toolbar button is touched or a gesture is recognized, and remains active only as long as the user maintains that particular contact with the screen. Modes are disengaged automatically when all fingers are lifted from the screen, providing an easy and reliable way to return to the default application state.

*Responsiveness:* Although this is only an early prototype, it was believed that the interface would need to be smooth and responsive in order for any user feedback to be meaningful. To give the user a real sense of direct manipulation,

objects would have to react without any hesitation or sluggishness. (That this intuition was correct was borne out in the comments from users who were all very impressed and pleased with the responsiveness of the application.)

## IV. PROTOTYPE APPLICATION

For the initial prototype of Flow, two versions were created, each employing a different user interaction style. The visual layout for both is essentially the same, featuring a toolbar along one side of the display and a large area for displaying the code being edited. (See Figures 1–3.) The toolbar contains buttons for creating each of the language elements (initially just three types of operations; input and output nodes; and datalinks for connecting nodes) as well as various commands (only Delete at the moment). The toolbar can be placed along either the left or right edge of the screen (for right- or left-handed users, respectively), as selected through a user preference.

In one version, the toolbar buttons control the current operating mode of the editor—adding an operation, a node, or a datalink, or deleting any of the above. A button must be “held down” (the touch must be maintained on the button) to engage the corresponding mode, thereby providing continuous, user-maintained kinetic feedback. (Figure 1.) While the mode is active, a second touch performs some manipulation within the context of the mode, such as tapping or dragging in the main editor pane. Although this can all be done (somewhat awkwardly) with one hand, the intent is for this interface to be used with two hands, the NP hand selecting the mode and the P performing the manipulation. This will be referred to as the “quasimodal” version, using Raskin’s term.

In the second version, all of the buttons for creating new objects are operated by dragging a finger from the button onto the code pane, which creates an object for the user to drag into place. (Figure 2.) The delete command is, for the moment, still operated using the two-handed method, but that could be replaced or supplemented with a gesture (such as making a stroke through an object, or dragging it to a trash area along another edge of the screen). This interface, despite the mixing of styles, will be referred to as the “drag-and-drop” version.

The drag-and-drop style should be familiar to most users from desktop GUIs, but the quasimodal interface may require a bit of explanation. It would initially be unfamiliar to most, but it was hoped that, once its operation is explained, it would be easy to understand and offer a viable alternative to more customary designs.

In both versions, the editor operates in a default mode when no buttons are pressed. In this mode, objects can be moved simply by dragging them (Figure 3). In the quasimodal version, objects cannot be moved while any of the buttons are engaged, as that action might be confused with a gesture related to the active quasimode. Take, for

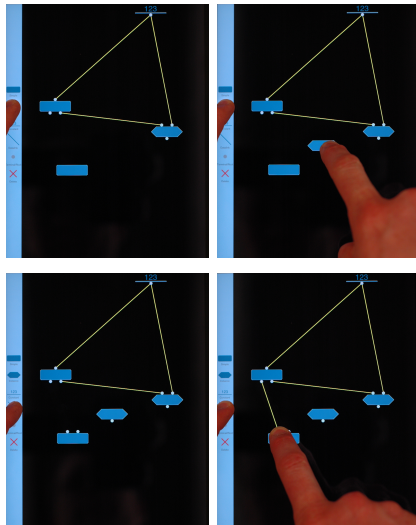


Figure 1. Quasimodal operation: Creating a new instance operation involves actively maintaining a touch on the Instance operation button (top-left photo) and tapping with another finger to place the operation icon in the code pane (top-right). To create a new datalink between nodes, the Datalink button is held (bottom-left) while another finger draws a line from the start node to the end node (bottom-right).

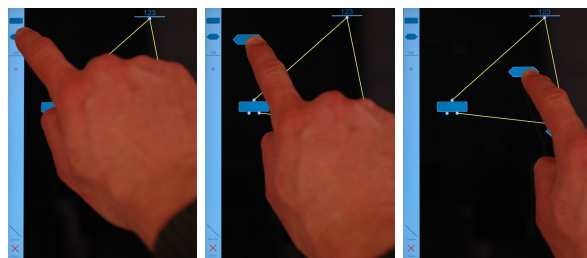


Figure 2. Drag-and-drop operation: New operations are dragged from the toolbar onto the editor pane and placed wherever desired. (Sequence shown left-to-right.)

example, creating a datalink between two nodes, which involves dragging one’s finger across the screen from one node to the other (while the Datalink quasimode is engaged). Nodes are attached to the top and bottom edges of operation icons, so the point of contact would be on or near the edge of the node’s operation icon. It would be hard to reliably distinguish this action from an attempt to move the operation, if it were not for the active quasimode providing context. Because a finger touch is not nearly as precise as a mouse or stylus, and some of the targets are necessarily small, the software must be forgiving and allow a good deal of leeway in hitting targets. By limiting the set of potential targets within a mode or quasimode, recognition accuracy can be much higher, which makes the interface seem more intelligent.

In the drag-and-drop interface, creating new code objects by dragging them from the toolbar invokes a kind of

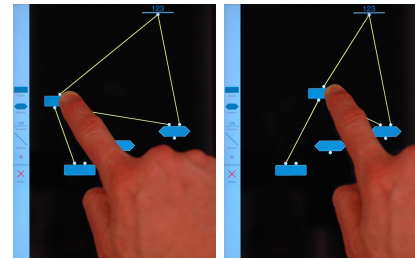


Figure 3. Moving objects: In the default editing mode (and at any time in the drag-and-drop interface), operation icons can be moved simply by dragging them.

quasimode as well, in as much as the creation mode is active only as long as the user maintains screen contact with the finger that started the drag gesture. However, unlike in the full quasimode interface, where engaging a toolbar button locks the entire interface into the corresponding quasimode, here the quasimode applies only to that one finger. Other fingers can continue to move existing objects in the editor, or begin additional creation actions. This feature is not actually inherent in the design; it was more an accident of some implementation choices, but it seems in the spirit of an immersive multi-touch interface to allow this to remain.

The target hardware for the prototype is a first-generation Apple iPad. The iPad has a 9.75" (247 mm diagonal) screen with a capacitive touchscreen overlay, which can simultaneously detect independent touches from all ten fingers. The software was developed as a “native app”, using Objective-C and Apple’s iOS SDK [11]. Videos demonstrating the prototype are available at [12].

## V. USER EVALUATION

A small sampling of users known to the authors were given the software to try, and were interviewed during and after using it. Some were expert Prograph users and so fully understood the model of the visual editor and the symbols used, while some others were non-programmers or novice Prograph users who, nonetheless, were able to give feedback on the responsiveness and naturalness of the interface options.

Overall, there was a strong consensus that the design was intuitive and very responsive. All but one user (and all of the expert users) felt that the sense of direct manipulation was greater than with a mouse, and that the responsiveness of the interface to touches was an important part of that.

The expert users all initially assumed that drag-and-drop would be the way to operate the toolbar buttons, while some novice users first tried to tap on the buttons to latch them on. They generally attributed this expectation to their previous computer experience. After an explanation was given (often as little as suggesting that one try using two hands), all but one quickly mastered the quasimodal operation, and the majority said that they preferred this version to using

drag-and-drop. Most found it faster to create new objects with the two-handed interface than by dragging items from the toolbar. And the expert users found adding nodes by dragging them from the toolbar to be somewhat awkward, perhaps because there was no analogue to this in the desktop version. No one had a problem adding nodes by tapping using the quasimodal version.

Two users expressed a preference for the drag-and-drop interface, primarily because they preferred to use just one hand. This seemed to be largely because the form factor of the tablet invites one to sit back and hold it in one hand, rather than placing it on a table so that both hands are free. One user suggested moving the toolbar buttons up towards the top edge of the screen so that the unit could be gripped up higher (better balancing the tablet in the hand) while still leaving the thumb free to press buttons. It was also suggested that a landscape orientation for the editor (it is currently portrait orientation only) would accommodate holding the tablet with both hands while operating the interface with one's thumbs.

## VI. CONCLUSION AND FUTURE WORK

The feedback obtained so far indicates that users find the environment appealing to use and that the sense of direct manipulation, a key feature of visual programming, is enhanced by the touchscreen interface. Having multiple, simultaneous touch points instead of single cursor opens up many possibilities for immersive and more efficient programming interfaces. While it is a bit early to begin judging coding efficiency, the experienced Prograph users were very excited and are looking forward to putting that question to the test with a future, more fully realized version.

However, it appears that users do need at least a small amount of instruction to figure out how to operate the quasimodal interface, given its unfamiliarity. Because mobile applications rarely come with any external documentation, some sort of in-program assistance should be added. Progressive enhancement techniques could also be used to gradually introduce novel interface features. Mobile games may offer a useful model to follow, as it is common for games to have specialized rules and control systems that must be learned before or during play. Testing so far does suggest, however, that once users learn the "trick", many do prefer the quasimodal interface, so the small amount of effort put into learning the unfamiliar system seems worthwhile.

Obviously, much more rigorous user testing is required, and we intend to conduct such testing as the prototype matures. Future work will largely be focused on experimenting with designs for navigating higher-level language constructs, such as methods, classes, and libraries. There are a number of general approaches to consider, such as zooming, overview+detail, focus+context (all described in [13]), and/or a more traditional hierarchical system. And within those approaches, there are several ways to map them to a

small screen and to take advantage of touch input. Again, particular attention will be paid to the option of bimanual input, with the hope that this can spur further investigation into the effectiveness of this mode of interaction.

## ACKNOWLEDGMENT

The authors would like to thank the members of the Dalhousie University Visual Languages and Design Group for their valuable input during the development of the Flow prototype, and their former Pictorius colleagues who took the time to provide expert-user feedback on the prototype.

## REFERENCES

- [1] P. T. Cox, F. R. Giles, and T. Pietrzykowski, "Prograph: a step towards liberating programming from textual conditioning," in *Proc. 1989 IEEE Workshop on Visual Languages*, 1989, pp. 150–156.
- [2] K. Yee, "Two-handed interaction on a tablet display," in *CHI '04 extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM, 2004, pp. 1493–1496.
- [3] M. Wu, C. Shen, K. Ryall, C. Forlines, and R. Balakrishnan, "Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces," in *Proc. 1st IEEE Int. Workshop on Horizontal Interactive Human-Computer Systems (TableTop 2006)*, 2006, pp. 185–192.
- [4] Y. Guiard, "Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model," *Journal of Motor Behavior*, vol. 19, pp. 486–517, 1987.
- [5] A. J. Sellen, G. P. Kurtenbach, and W. A. S. Buxton, "The prevention of mode errors through sensory feedback," *Human-Computer Interaction*, vol. 7, pp. 141–164, Jun. 1992.
- [6] J. Raskin, *The Humane Interface: New directions for designing interactive systems*. Reading, MA, USA: Addison-Wesley, 2000.
- [7] M. Nielsen, M. Störring, T. B. Moeslund, and E. Granum, "A procedure for developing intuitive and ergonomic gesture interfaces for HCI," in *Gesture-Based Communication in Human-Computer Interaction*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2004, vol. 2915, pp. 105–106.
- [8] D. Mauney, J. Howarth, A. Wirtanen, and M. Capra, "Cultural similarities and differences in user-defined gestures for touchscreen user interfaces," in *Proc. 28th Int. Conf. Human Factors in Computing Systems (extended abstracts) (CHI EA '10)*. New York, NY, USA: ACM, 2010, pp. 4015–4020.
- [9] M. Frisch, J. Heydekorn, and R. Dachsel, "Diagram editing on interactive displays using multi-touch and pen gestures," in *Diagrammatic Representation and Inference*, ser. Lecture Notes in Computer Science, A. Goel, M. Jamnik, and N. Narayanan, Eds. Springer Berlin / Heidelberg, 2010, vol. 6170, pp. 182–196.
- [10] D. A. Norman and J. Nielsen, "Gestural interfaces: a step backward in usability," *interactions*, vol. 17, pp. 46–49, Sep. 2010.
- [11] Apple Computer. (2011) iOS Reference Library. [Online]. Available: <http://developer.apple.com/library/ios/navigation/>
- [12] M. Hackett, "Multitouch dataflow editing (video)," 2011. [Online]. Available: <http://web.cs.dal.ca/~pcox/visual/Multitouch/>
- [13] A. Cockburn, A. Karlson, and B. B. Bederson, "A review of overview+detail, zooming, and focus+context interfaces," *ACM Computing Surveys (CSUR)*, vol. 41, p. 2:1–2:31, Jan. 2009.

# Design Guidelines for Hybrid 2D/3D User Interfaces on Tablet Devices

## A User Experience Evaluation

Katri Salo

Intel and Nokia Joint Innovation Center  
 Department of Information Processing Science  
 University of Oulu, Finland  
 katri.p.salo@gmail.com

Leena Arhippainen, Seamus Hickey

Intel and Nokia Joint Innovation Center  
 Center for Internet Excellence  
 University of Oulu, Finland  
 leena.arhippainen@cie.fi, seamus.hickey@cie.fi

**Abstract**— There is a growing proliferation of 3D based applications in tablet devices, but there is a lack of studies evaluating user experiences of these user interfaces. In particular, most of these applications use a mix of overlaid 2D controls and embedded 3D controls for user interactions and there is little current understanding on how users perceive and experience these controls. This paper presents the results of a user experience evaluation made for the user interface of four 3D applications running on two different tablet devices. A number of results are obtained. First, the results show that users wish that applications had less 2D overlaid icons and more space for touch gesture interactions. Second, the number of on screen control elements complicates the activity and provides a more disappointing experience. Third, avatar control was difficult as there were no tips or clear cues on how to use them. Fourth, users expected to control and interact with the applications by using direct touch gestures in the 3D space. As a result, design guidelines for hybrid 2D/3D user interfaces were created.

**Keywords**-3D; hybrid; tablet; user experience; user interface

### I. INTRODUCTION

There is an increasing number of products and services which are available for mobile tablet devices. In particular, there has been an increasing proliferation of three-dimensional (3D) graphical-based applications, which have been developed for tablet devices. Dillon and Morris [1] state that users are often unwilling to use technology which would result in an unimpressive performance and this is particularly relevant to 3D applications which add complexity to the User Interface (UI). From the user's perspective, there needs to be more research in the touch screen field [2] and in the 3D implementation area [3][4].

This paper presents the results of a User eXperience (UX) study of existing 3D applications that are commercially available on tablet devices such as the Apple iPad and Samsung Galaxy Tab. We studied the user experience of four 3D applications, three games and one map application. The focus was on the UI controls and not on the games' playability. Of these, the games make heavy use of an overlaid UI with 2D icons, as well as touch gestures and embedded objects in the 3D space. Each application takes a different approach to solve similar problems and this study evaluates these choices from the user experience point of view and identifies the principal design challenges.

This paper is organized as follows. Section II presents research that is relevant to this research topic. Section III presents the research methods used. Section IV presents the results. Section V presents a discussion of the results. Section VI concludes the research paper.

### II. RELATED WORK

In 1974, Elographics developed and patented the five-wire resistive technology, the most popular touch screen technology, which is still in use today [5]. In 1982, the first multi-touch system was developed at the University of Toronto; this system allowed users more than one contact point at a time. Gestural interfaces have become popular for the mass market during the past decade. In 2007, Apple introduced the iPhone and iPod Touch demonstrating their touch screen capabilities. [5]

Yee [6] summarizes five major criteria which are thought to contribute to the effectiveness of gestural interactions. First, the application or system interface should make it clear to the user that gestures can be used. Second, the gestures should be obvious and intuitive for the relevant tasks. Third, the users need to be allowed to gesture with minimal effort, e.g., interaction is simple to perform and without a need for unusual motor skills. Fourth, gestures should have a logical relationship with the application functionality that they represent, i.e., in the type of movement and type of interaction with objects that they are acting on. Fifth, gestures should be designed for repetitive use and minimal muscle stress. [6]

With touch screen gestures it is easier for users to carry out certain actions, depending on the task they want to accomplish [7]. It has been recognized that the aesthetical appeal of the interface is important for users [8] and it can also improve the user performance.

There is little to be found in literature on hybrid UIs; the available literature is more focused on virtual environments. Haan et al. [9] describe hybrid UIs as an incorporation of 2D UI elements in a 3D environment. This type of solution is often used in games. In a hybrid UI, the elements in use should not be unreasonably large, because they are intended as supplements rather than the main focus. The interface should be intuitive to use. Icons should provide tool tips so that it is also easy for inexperienced users to interact with. The icons should be provided with a symbol instead of text to save space for the embedded UI. [9]



Yoon et al. [4] compared a conventional 2D image-based system and a web 3D system by using an online survey of furniture style preferences. The findings showed 3D to be superior to 2D for product examination and decision making.

3D is expected to emerge from films to personal consumer electronics. 3D can provide a greater immersion experience for users, but only if this added value is not restrained by poor image quality. According to Jumisko-Pyykkö et al. [10], an overall excellence of 3D is also influenced by the quality of the display. The final quality is determined by the users' perception, which is influenced by their characteristics and the context of use. [10][3]

### III. RESEARCH METHOD

The use of hybrid 2D/3D UIs were studied by conducting user experience evaluations for four different 3D applications. Three were three games and one was a 3D map application. The selected games were Order and Chaos (O&C), Pocket Legend (PL) and Dungeon Defenders (DD), which are all mobile multiplayer online (MMO) games. These were selected because they have the same game objectives but different approaches to the user interface. The map application was the YesCitz Barcelona (YB) demo version. The applications have an overlay UI, which is the layer where control widgets and other functions with 2D icons are found. The 3D space has an embedded UI with 3D objects and models. O&C and YB run on an Apple iPad, and PL and DD run on a Samsung Galaxy Tab. We selected two different tablets in order to observe how people hold the device and how they interact with the touch screen. For instance, we are particularly interested in which hand or fingers they use for holding and controlling the device. This paper discusses these issues when they are related to the viewpoint controls and both 2D and 3D embedded object interaction.

#### A. Test Setting and Methods

For studying how users experience the interaction with hybrid UIs, we used various methods; observation, interviews, user tests, a customized version of the product reaction cards [12] method and a discussion forum. After using each test application, the user was interviewed by using the product reaction card method. They selected five suitable adverbs from a list of 39, and gave a short argument for their selection. This provided information that either supported the observation or brought new findings.

In the evaluation, participants used each application for around 10 minutes. During the use, no detailed tasks to accomplish were given to participants. However, users were asked to move the avatar, familiarize themselves with the game environment and interact with objects (other players, enemies, etc.) when they appear. Users were asked to think aloud and reply to questions concerning the use of applications and the tablet device. If there were problems in proceeding or performing a certain function, they were given tips. The observation focused on the touch gestures, the use of the overlay and embedded UI objects and also the hand and finger positions while holding the tablet and controlling the application.

Before the actual tests, a pilot test was performed to identify potential problems, identify gaps and ensure timely evaluation completion. A pilot test was carried out and changes were made to the product reaction card method and the order in which the applications were shown to the users. It was decided to alternate the order of the applications in the future. O&C was the first game presented to eight users. This is because of different UI solution: the use of a hidden control widget. User evaluations were conducted in laboratory settings. The duration of the evaluations varied between 1-1,5 hours. All sessions were video recorded.

#### B. Participants

We had 12 participants, whose age varied from 23 to 34 years, while the median age was 28 years. The ratio between males and females was 1:1. The age group for the evaluation was selected based on the target user group, according to tablet and mobile gaming demographics [13][14]. One participant was left handed, and the rest were right handed. Seven users had prior experience with touch screen mobile devices while five had none. Only one participant actively used a tablet device. Eleven participants had prior experience with 3D technology, for example from films or games, but one had not. When participants described their gaming experience, eight turned out to be active gamers on several different platforms. The remaining four had no experience or they had used rather dated games a long time ago.

### IV. RESULTS

This chapter presents the results on each test application and tablet. Sections A and B discuss the map and avatar control. Section C describes the user interaction with embedded UI objects, and section D discusses switching between the overlay UI and the embedded UI.

#### A. Map Controlling

YB differs from the game applications by only having a single UI overlaid component which just changes the camera orientation to 'north'. The interaction is based on direct gestures to the embedded UI. The application opens with a bird's eye view over the area, and the user must then zoom closer to street view, where the environment is modeled in 3D (Fig. 1). A pan gesture is needed to adjust the horizontal view and it was the most problematic for users due to its unfamiliarity. Six different touch gestures were tried. In addition, one user tilted the device in the air to pan the view, but this kind of feature was not possible. A few participants commented on the similarity of the rotate and pan gestures. The 3D modeling received positive comments and it was considered helpful. There were participants, who first tried to find the 2D control widget for controlling the UI, but eventually they understood that the application does not have any widget for zooming, rotating or panning. These participants had tested one or more of the games before the map application, so we suspect that they probably learned that 2D elements are used for interaction. One participant with no game experience said: "... the map looks so neat-maybe that's why I was more eager to experiment with gestures compared to the games."



Figure 1. YB: Pan gesture used (right hand) to tilt in street view.

B. Avatar Control

Avatar control was necessary in three of the applications, O&C, PL and DD. Table I shows the order (O) in which the applications were used by each user, the amount of time (S) they spent trying to locate the avatar control widget and whether they had game experience (GE) or not (NE). The order of the tested applications is important, because previous usage has influences on the expectations and assumptions for the next one. O&C was the first game in 8 tests; PL was the second game in the 8 tests and DD as the third game in 6 tests. (Table I)

1) Hidden Avatar Control in O&C

The avatar control widget in O&C is located in the bottom left corner and is initially invisible until the user touches that area, at which point it becomes visible and controllable (Fig. 2). This was the most difficult gesture for the participants to figure out. In the beginning, most participants assumed that the avatar would be controlled by either tapping or sweeping the game space, or tapping an icon with a 'boot' symbol. Many said that a touch screen device mislead them into believing that the avatar is controlled in a similar manner as with a mouse and a keyboard-like point and click, or by a gesture in the game space. When participants opted to use their right hand for controlling the avatar, this blocked a lot of the view on screen (Fig. 2). They wondered why the widget could not be activated from the right side of the UI. A participant tried to solve the problem by flipping the tablet upside down, thinking that the UI would adapt differently. There were mainly negative comments on the use of an invisible widget as the participants did not feel it was necessary to hide it. In total, there were 11 different gestures that the participants tried to control the avatar, but in average they needed to try five gestures before finding the correct one. It took in average 67 seconds to find the correct gestures (Table I).

After getting used to the hidden control widget in O&C, some users eventually gave positive comments on the unrestricted control area, which is demonstrated in Fig. 2. In O&C it is also possible to have the avatar move automatically to a new position by clicking on an object with a special marker. This marker comes in the form of a 2D yellow arrow icon, which momentarily appears on the

TABLE I. SECONDS (S) FOR FINDING AVATAR CONTROL WIDGET AND THE ORDER (O) OF THE APPLICATION USE IN THE TEST. 8 USERS HAD PRIOR GAME EXPERIENCE (GE) AND 4 HAD NOT (NE)

GE/NE	Order & Chaos (O&C)		Pocket Legends (PL)		Dungeon Defenders (DD)	
	S	O	S	O	S	O
1GE	48	1	41	2	6	3
2GE	28	1	3	2	9	3
3GE	5	3	7	1	9	2
4NE	23	3	175	2	51	1
5NE	56	1	2	2	7	3
6GE	31	3	58	2	57	1
7NE	100	1	378	2	10	3
8GE	152	2	8	3	70	1
9GE	136	1	36	3	31	2
10NE	116	1	64	2	14	3
11GE	70	1	31	2	6	3
12GE	35	1	28	3	35	2
Average S	67		69		25	



Figure 2. O&C: Avatar control widget dragged through the game space.



Figure 3. PL: Control widget (left) is used with press and drag gesture.

embedded UI. None used this. One reason could be that it went unnoticed, since it typically appears on the bottom of the UI, where the other hand was using the camera control the overlay UI icons. It is possible that this arrow can be understood as a guide to a location as well.

2) Overlaid and Embedded Avatar Control in PL

To control the avatar in PL, the control widget has a symbol of a yellow, eight-pointed star. When pressed, the widget activates a yellow circle around it (Fig. 3). During the test, participants commented that this does not remind them of a control widget, but rather that it looks like a compass. Five different gestures were used. In average, it took three

gestures and 69 seconds (Table I) to find it, one reason being the unclear symbol. The other reason is that it is possible to control the avatar navigation by tapping the game space which causes a 'cross' indicator to appear and the avatar walks towards it. Three participants found this in-space tapping mechanism first, before noticing the control widget. Four participants never found this control option for tapping the space. For the eight users who did eventually find this in-space tapping option, it took an average of 144 seconds to notice it. These eight persons considered the in-space tapping action as a preferred choice to using the control widget. They stated that it was easy to use their thumb for this in-space navigation and this choice was affected by the smaller screen size of the Samsung Galaxy Tab.

### 3) Overlaid Avatar Control in DD

The DD avatar is controlled by a grey widget on the bottom left corner (Fig. 4), which was found rather quickly by all participants. The average time to find this widget was the shortest of all the games, 25 seconds (Table I). Participants experimented with different gestures to control the avatar. In total, five different types of gestures were tried, of which one is the correct gesture. In average, each user tried two incorrect gestures before finding the correct gesture to control the avatar. Tapping the game space was tried by eight, sweeping the game space by seven. A direct tap on the avatar was used three times and a tap for the overlay UI icons five times. Users felt that the symbol for the control widget was quite understandable (Fig. 5). Still many of them expected to have a possibility to control the avatar by touching the game space directly.

### C. User Interaction with Embedded UI objects

In O&C, there are 2D icons such as talk bubbles or message scrolls appearing above the embedded UI objects to indicate interaction. If there was this kind of a cue, participants mainly tapped them, instead of tapping on the object (e.g., avatar) directly. A direct tap on an object is possible in this game, and three participants also tapped on objects (e.g., doors) which did not have any function. Another cue used in O&C for indicating interactable objects are colored circles below the avatar or enemy, which was easy to notice, and participants were then able to interact directly with them. It was observed that using two hands simultaneously was not comfortable for everyone.

In PL, a bow icon in the overlay UI is quite large and noticeable compared to the size of other icons on the overlay UI. It was tested by three participants in an attempt to interact with an embedded UI object. However, the two correct gestures were quickly discovered. Six participants were running through portals with a text "magic portal" above them; eight participants used the tap on the embedded object.

For the participants, the interaction mechanism with embedded UI objects in DD was confusing. There are three possible ways to interact, depending on the object. Also there are inactive objects with no functionality, such as doors or tables, yet two participants tapped these types of objects. Some objects such as caskets can be bumped into to open



Figure 4. DD: A tap used to interact with an embedded 3D object (under right thumb).



Figure 5. DD: A 2D icon (under right thumb) used to interact with a 3D object on game space.

them, and this was used by three users. The most used gesture was to tap an object directly, which was eventually done by ten users. There is an option to interact with a tap on a corresponding 2D icon, which appears momentarily on the overlay UI. This was used by six participants. Some of them had already tried to tap the embedded object, but nothing happened due to the distance from the object, or the touch pressure was too light. From then on, they continued to interact through the corresponding 2D icon. When participants were given a tip to find a certain object and then to try interacting with it, four expected it to happen by using a sword icon and five tapped a menu icon. One participant tapped a blue arrow icon.

### D. Switching Between Overlay and Embedded Objects

According to participants, the icons were mainly "misleading symbols". A boot symbol in O&C was the most problematic; it was interpreted as a kick, walk and weapon to name a few. Also, the amount of icons was important; too many icons caused user confusion. The result was that the users focused more on the overlay UI icons than using what is available in the embedded UI. In an interview, it was remarked that: "... if this is a touch screen device, why do you need icons to do stuff you could use gestures for?"

In the embedded UIs, there are different cues (color, text) to inform the gamer that one can interact with an object. Eleven participants playing O&C did not need a suggestion to try interacting with the game space. The O&C UI was also

easy to learn and more intuitive. The objects are often animated or color cued (enemies, avatars), or cued with a 2D symbol above the embedded object. This raises a question on whether the choice of having 2D icons in the embedded 3D UI is consistent with the idea of a 3D space. It seemed that participants were more prone to interact through these 2D icons whenever they were available. It could be possible that since O&C was the game which has the least number of overlay UI icons, more interaction happened on the embedded UI.

Even though PL has the most overlay UI icons, seventeen of them, the use of the embedded UI was active and quickly picked up. Since PL has the chance to control the avatar by tapping on the game space, it may have encouraged eight users to interact with objects. Of the overlay 2D icons, the icon with a large bow symbol was tried by three users. Also, the cues were simple, mainly text and animation.

DD has 14 overlay UI icons, and when there is a possibility to interact with an object, a 2D icon appears on the overlay UI (Fig. 5). So the gamer can choose between the icon and the object directly. Six users were interacting with the embedded UI objects (Fig. 4) directly without help, the remaining six were either encouraged to find a way to do that or they never noticed the possibility. The different overlay icons tapped for interaction had symbols of a blue arrow, two swords and a blue menu. It was a common error that a participant did not react in any way to an animated, glowing large white stone with text on it (Fig. 4). It is a portal to the next level, yet they passed it multiple times until they were advised to tap it.

## V. DISCUSSION

Participants expected to use the touch gestures instead of interacting with multiple 2D icons. Even though Chehimi and Coulton [11] mention that mobile games should be playable with a single button, it is not always possible, especially in the MMO genre. Still the UI should be simplistic to give more room and chances for the touch gesture controls. It was apparent that the test applications did not meet Yee's [6] recommendation to design interaction for repetitive use and minimal muscle stress. Many participants felt that the control widget use was tiring and had problems to operate with it. A similar symptom was however not apparent with YB use, where the user has greater freedom to perform gestures.

In all the game applications, participants expected that the controls would happen with a direct touch on the game space, usually by tapping or swiping it. A few of the comments concerned the limited area for the touch, that when using a fixed control widget, one has to monitor where their fingers are. No one considered this input method particularly advanced. In PL, there is the possibility to use the tap gesture on the game space, which was preferred over the fixed control widget due to the unrestricted touch area and also because both hands could be in use.

Another issue was the typical location of the control widget, which was on the left in each application. The majority of the participants were right handed, and often held the tablet with their left hand, controlling with the right hand.

This blocked the view on the game space, and caused another complaint. When it was suggested that they try controlling with their left hand, some felt that they did not trust the strength of their thumb to press and drag. The users then tried to control with their left thumb, but they soon reverted back to their left or right index finger. Most of the participants expected that it would be possible to change the control widget to the right side of the UI.

The biggest problem in games according to participants was the avatar control; there were no quick cues on how – to control the avatar movement. A brief animation of the gestures during the first steps of the application use could solve this. One commented that: “... if a game is supposed to be played with two thumbs, then there shouldn't be important objects in the middle of the screen, you can't reach them.” Similar observations were made during the tests when some female participants had problems with the Apple iPad and the simultaneous use of two hands.

### A. Proposed Design Guidelines

Based on the findings from the user evaluations, we propose the following set of design guidelines to improve the user experience of the hybrid 2D/3D user interfaces in tablet devices:

- 1) *Controls*
  - Place an easily identifiable control widget in a logical location on the overlay UI
  - Promote the use of direct on-screen touch gestures
  - Provide quick and interesting guidance for the principal gesture use.
  - Reduce the need for an overlay UI control use
  - Avoid implementing gestures which are too similar
- 2) *3D space*
  - Design clear and consistent cues for interactive embedded 3D objects
  - Avoid use of 2D icons in 3D space
  - Design location of interactive elements by taking into account user's hands and fingers positions
- 3) *Icons*
  - Give the user a possibility to customize the icons' location
  - Use simple and consistent icons in the overlay UI
  - Minimize the amount of icons in the overlay UI
  - Ensure that the size of the icons in the overlay UI is consistent
  - Avoid placing icons too close to each other.

These guidelines are derived from the results of the study of the use of hybrid 2D/3D UIs. These can be applied when there is a need to use both 2D icons in the overlay UI and 3D objects embedded into space.

## VI. CONCLUSION AND FUTURE WORK

A user experience study on 3D applications in tablet devices was conducted. Twelve participants used four applications in two tablet devices. The study found issues in a combined use of an overlay UI and an embedded UI,

focusing in the control methods and interaction with embedded objects in a 3D space.

2D icons on the overlay UI negatively affect the use of direct gestures in 3D space. These overlay UI elements often result in an unsatisfying interaction, and reduce the user's interest on embedded objects in 3D space. Availability of 2D icons also distracts the user from interacting with embedded 3D objects. An application without a forced use of overlay UI controls gets users more interested in experimenting with the touch gestures and the 3D space.

Based on the study results, design guidelines for a hybrid UI was created with an emphasis on encouraging direct on-screen gestures and reducing the dominance of an overlay UI. In the future with 3D UI design, there might be a need to design only embedded UIs with 3D objects and totally avoid 2D overlay icons. Future studies will be conducted to validate these guidelines.

#### ACKNOWLEDGMENT

This work was carried out at the Intel and Nokia Joint Innovation Center as part of Tekes Chiru project. We want to thank Lotta Haukipuro from CIE for an excellent co-operation with the PATIO test user recruitment. We would like to warmly thank all the participants and Terhi Koskipaasi for assisting with experiments.

#### REFERENCES

[1] A. Dillon and M. Morris, "User acceptance of new information technology: theories and models," in Annual Review of Information Science and Technology, vol. 31, M. Williams (ed.), Medford, NJ: Information Today, 1996, pp. 3-32.

[2] A.A. Ozok, D. Benson, J. Chakraborty and A.F. Norcio, "A Comparative Study Between Tablet and Laptop PCs: User Satisfaction and Preferences," International Journal of Human-Computer Interaction, vol 24, Mar. 2008, pp. 329-352, doi:10.1080/10447310801920524.

[3] S. Jumisko-Pyykkö, M. Weitzel and D. Strohmeier, "Designing for User Experience: What to Expect from Mobile 3D TV and Video?" Proc. of the 1st International Conference on Designing Interactive User Experiences for TV and Video (UXTV '08), ACM Press, Oct. 2008, pp. 183-192, doi:10.1145/1453805.1453841.

[4] S. Yoon, J. Laffey and H. Oh, "Understanding Usability and User Experience of Web – based 3D Graphics Technology," International Journal of Human – Computer Interaction, vol 24, Mar. 2008, pp. 288-306, doi: 10.1080/10447310801920516.

[5] D. Saffer, Designing Gestural Interfaces. Sebastopol, CA: O'Reilly Media Inc., 2009.

[6] W. Yee, "Potential Limitations of Multi-touch Gesture Vocabulary: Differentiation, Adoption, Fatigue," Proc. of the 13th International Conference on Human-Computer Interaction (HCI 2009). Part II: Novel Interaction Methods and Techniques, Springer Berlin / Heidelberg, Jul. 2009, pp. 291- 300, doi: 10.1007/978-3-642-02577-8\_32.

[7] F. Balagtas-Fernandez, J. Forrai and H. Hussmann, "Evaluation of User Interface Design and Input Methods for Applications on Mobile Touch Screen Devices," INTERACT '09: Proc. of the 12th IFIP TC 13 International Conference on Human-Computer: Part I, Springer Berlin / Heidelberg, Aug. 2009, pp. 243-246, doi: 10.1007/978-3-642-03655-2\_30.

[8] S.J.P. McDougall and I. Reppa, "Why do I like it? The Relationships Between Icon Characteristics, User Performance and Aesthetic Appeal," Proc. of the Human Factors and Ergonomics Society 52nd Annual Meeting, Sep. 2008, pp. 1257-1261.

[9] G. Haan, E.J. Griffith, M. Koutek and F.H. Post, "Hybrid Interfaces in VEs: Intent and Interaction," Proc. of the 12th Eurographics Symposium on Virtual Environments (EGVE 2006), May 2006, pp. 109-118, doi:10.2312/EGVE/EGVE06/109-118.

[10] S. Jumisko-Pyykkö, D. Strohmeier, T. Utriainen, and K. Kunze, "Descriptive Quality of Experience for Mobile 3D Video," Proc. of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordCHI '10), ACM Press, Oct. 2010, pp. 266-275, doi:10.1145/1868914.1868947.

[11] F. Chehimi and P. Coulton, "Motion Controlled Mobile 3D Multiplayer Gaming," Proc. of the 2008 International Conference on Advances in Computer Entertainment Technology (ACE '08), ACM Press, Dec. 2008, pp. 267-270, doi: 10.1145/1501750.1501813.

[12] C. Barnum and L. Palmer, "More Than a Feeling: Understanding the Desirability Factor in User Experience," Proc. of the 28th International Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '10), Apr. 2010, pp. 4703-4715, doi: 10.1145/1753846.1754217.

[13] K. Crosett, "Mobile Game Marketing to Increase", 18.3.2011, Available: <http://www.marketingforecast.com/archives/10608>

[14] Marketing Charts, "iPad, iPhone Users Trend Younger, Wealthier" 2010, <http://www.marketingcharts.com/direct/ipad-iphone-users-trend-younger-wealthier-13778/nielsen-ipad-user-profile-august-2010.jpg>.

# A Usage-Centered Evaluation Methodology for Unmanned Ground Vehicles

Jurriaan van Diggelen, Rosemarijn Looije, Tina Mioch, Mark A. Neerincx, Nanja J. J. M. Smets  
TNO

Kampweg 5, 3769 DE Soesterberg, The Netherlands

{jurriaan.vandiggelen, rosemarijn.looije, tina.mioch, mark.neerincx, nanja.smets}@tno.nl

**Abstract**—This paper presents a usage-centered evaluation method to assess the capabilities of a particular Unmanned Ground Vehicle (UGV) for establishing the operational goals. The method includes a test battery consisting of basic tasks (e.g., slalom, funnel driving, object detection). Tests can be of different levels of abstraction, and be performed in a virtual or real environment. In this way, several candidate UGV's in a procurement program can be assessed, and thus compared. Also, it can give directions to research on improving human-robot interfaces. A first case study of this methodology conveyed capability differences of two alternative user interfaces for a specific UGV with their operational impact.

**Keywords**—Human-robot cooperation; Performance evaluation

## I. INTRODUCTION

Usage-centered Unmanned Ground Vehicle (UGV) evaluation is important for a number of purposes. It can be used in a procurement program, allowing several off-the-shelf candidate UGV's to be assessed against their operational needs, and thus compared. Furthermore, it can give directions to UGV development by clearly identifying shortcomings in operator-robot interaction. It can also help to select an adequate set of components (e.g., sensors, controllers, user interfaces) to be combined in the robot.

Nevertheless, performing a structured, well-founded UGV evaluation in practice poses a number of difficulties. Firstly, the usage environment may not be easily accessible. This can be because the UGV is intended to be used at a remote or dangerous location, or because the operator's final work environment is difficult to simulate in a laboratory setting. Secondly, the UGV may not be entirely available at the time of evaluation. For example, the UGV platform may be available and ready for testing before a decision is made about which sensors will be mounted on the UGV. Another example occurs when a UGV is still in its specification phase, and the manufacturers wish to perform an early evaluation of the requirements baseline before actually buying the physical hardware.

Because current UGV benchmarks are scarce and fail to adequately address these problems, we have developed a usage-centered evaluation methodology. The methodology is based on a test-battery and makes no prior assumptions on the *location* of the evaluation, or the UGV's *phase of development*. A test can be regarded as an *exam exercise*, constituting a simple atomic task, such as doing

slalom, funnel driving, or object detection. If the robot, controlled by a qualified operator, passes the exam, it can be concluded that it has the basic capabilities that were identified as critical for the operational task. Because tests can be defined in an abstract way, and can be performed in both the real world as well as in the virtual world, they do not impose any constraints regarding the location of the UGV evaluation. Because tests are designed to be as elementary as possible, they can be used to partially evaluate a robot which has not been fully assembled yet. As missions are changing substantially, robot technology is progressing quickly, and relevant human factors knowledge is increasing continuously [1], the test battery should not be seen as a static entity. Rather, we regard it as a standard toolkit for UGV operators which should be updated regularly and tailored to the specific situations encountered. Furthermore, an additional summative evaluation with adequate fidelity will most often be needed at some point in the development cycle to assess all the dependencies between context, work organization, personnel, UGV and operational outcomes.

The purpose of this paper is threefold. Firstly, we present a structured and well-founded methodology for usage-centered UGV evaluation and design. Secondly, we give guidelines for UGV test development, considering issues such as abstractness and realism of tests and how this affects their validity, taking into account operational demands, human factors issues, and technological constraints and opportunities. Thirdly, we present a standard test battery and report on our experiences on using the methodology for robot evaluation. For a more in-depth investigation of the soundness of our approach, the reader is referred to [2].

We have started developing the standard test battery by designing around forty basic abstract tests together with the corresponding evaluation criteria, and their relation to operational demands. Because the tests are abstract, we can perform these tests at the location of our customers (either procurement officers or UGV manufacturers). The current test battery contains tests which must be performed in the real world, i.e. we have not yet included any virtual environment tests. Nevertheless, we have set up the methodology in such a way that we can straightforwardly extend the test battery with tests in a virtual environment, e.g., USARSim [3]. This allows us to design tests with virtual robot configurations in simulated environments, which are

hardly available for (cost-effective) robot evaluations and/or might involve a danger of damaging the robot.

The paper is organized as follows. In the next section, we describe our abstract evaluation framework, and how testing can be performed at different phases of development. In Section III, we describe how tests can best be designed. Section IV describes an initial version of our basic test battery, and reports a case study, followed by a conclusion in Section V.

## II. UGV EVALUATION METHOD

The methodology follows a human-centered approach, i.e., we focus on a human operator who interacts with a UGV within a certain context. This is depicted in Figure II.

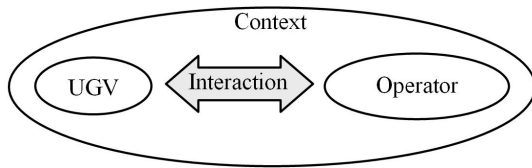


Figure 1. UGV-Operator System in Context.

Each of the four components in Figure 1 has an influence on the overall performance of the system. Below, we will outline some relevant factors for each of these components.

- UGV factors: These factors are typically described in the UGV’s specification document. Examples are the energy consumption properties of a UGV, the amount of horsepower, the availability of different sensors, the availability of robotic arms and grippers, the maximum speed, size, weight, etc.
- Interaction factors: These are factors concerning the interaction between the UGV and operator. This may concern information which must flow from UGV to operator, such as sensor images and information on the slope of the terrain. It may also concern information which flows from operator to UGV, such as directions to the UGV that it must adjust its camera angle, or any other type of control action performed by the operator.
- Operator factors: Examples are: knowledge, skills, abilities and training level of the operator, fatigue, motivation, etc.
- Context factors: Example of relevant context factors are: properties of the soil, weather conditions, light conditions, etc.

Obviously, these four factors are interrelated. For example, darkness (a context factor) may obstruct proper UGV operation, unless the UGV has a night vision camera (UGV factor), and the interface allows to properly view these camera images (an interaction requirement). In this paper, we have chosen not to evaluate these different aspects separately, but to take all of these aspects into account at once. Hence, the evaluation measures the total operator-UGV performance.

### A. Situated Cognitive Engineering

Our UGV evaluation methodology is a special component of situated cognitive engineering [4] that views system development as an iterative process in which system’s functions are specified and assessed in a systematic way to establish a sound foundation of the specified and/or selected functions. An overview of this evaluation methodology is depicted in Figure 2.

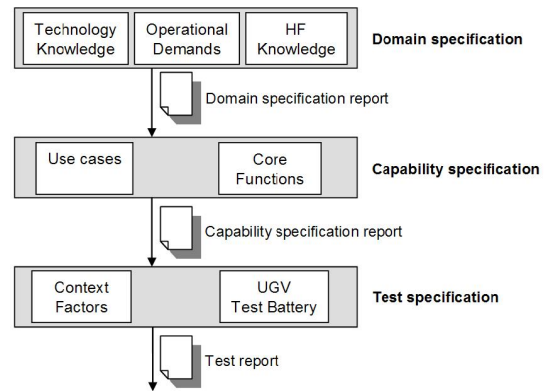


Figure 2. UGV Evaluation Process.

In the domain specification layer, the technology knowledge, operational demands and human factors knowledge are investigated. Usually this is done in a series of workshops with HRI experts and envisioned end users. During these workshops, the HRI experts are provided with information on what kinds of operations the users intend to use the UGV for, and what kind of UGV candidates they were considering. This information results in the domain specification report which specifies the following aspects:

- Envisioned operational use (which kind of environments, on which terrains, indoor/outdoor, weather types, envisioned operators, etc.)
- Initial technological investigation (potential UGV candidates, using which sensors and which interfaces, potential technological pitfalls)
- A list of human factors issues which are relevant for the envisioned technology and operational demands (such as operator sickness or risk of information overload).

The domain specification report forms the basis upon which the use cases and core functions are derived, as depicted in the capability specification layer. These two components are described in the capability specification report, which makes the required functionality of the robot operator system more concrete. This report describes a list of use cases which demonstrate the nominal and extreme use of the UGV, and a list of core functions.

In the final phase (the test specification layer), this knowledge is further refined into a concrete list of relevant context factors, and a concrete UGV test battery to which

the robot is subjected. These tests comprise assessments for capabilities that are tailored to the (envisioned) UGV-supported operations, technological demands and human factors. The final judgment whether the robot is appropriate, is reported in the test report which describes the following:

- A selection of tests from the test battery, including information on why the test was selected, how the test was instantiated (which context factors were taken into account, practical constraints, setup of the environment).
- The UGV-operator system performance on the different tests, describing the performance on objective measures and subjective measures (feedback from the user about their performance).
- A final judgment which summarizes the test results.

The seven components represented by the white boxes in Figure 2 can be regarded as an evolving toolbox, which grows over time when more UGV's are subjected to the evaluation method and more operational demands and human factors are taken into account. The advantage for the UGV-evaluator is that he does not start from scratch each time a UGV is evaluated; the advantage for the method is that it affords continuous updates of the task battery. Thus, the evaluator can reuse use cases, tests, technology knowledge, and so forth, which have been developed in previous UGV evaluations. In this way, a set of generic, reusable core components are iteratively developed.

For a specific evaluation, an instance of the core test battery should be formulated, consisting of the selection and prioritization of the tasks and criteria for evaluation. Because the foundations of the core test battery evolve continuously, the compilation and definition of the constituting tasks should be updated regularly. As depicted in Figure 2, the three evaluation reports form the milestones in this process.

*B. Evaluation dimensions*

In general, evaluation experiments can differ in fidelity and realism [5]. We can apply the same categorization to the UGV evaluation tests. In this domain, fidelity indicates how close the test environment resembles the environment in which the UGV is planned to be used. For example, we can perform low fidelity tests in a laboratory, or in the real environment (high fidelity). Realism varies from one extreme-the real environment-to the other, a virtual environment. For example, instead of going to a real earthquake site to test a prototype, the prototype can be tested in a virtual environment. The test space is depicted in Figure 3.

Typically, UGV's are evaluated in a series of tests, starting with easily performable tests with low fidelity and realism to experiments in the real world. This is indicated by the arrows in Figure 3.

An example of a low fidelity and low realism test is a cognitive walk through. This was for example done in the Mission Execution Crew Assistant (MECA) project [4],

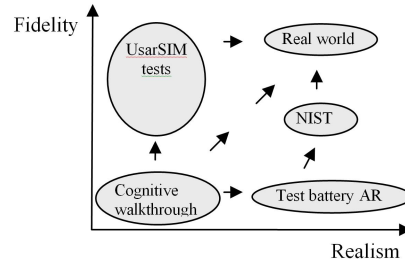


Figure 3. UGV Test Space.

which uses Situated Cognitive Engineering to validate the requirements baseline of an astronaut's cognitive support system. During the cognitive walkthrough evaluation, participants were talked through a use case and answered questions during and after the walkthrough. Storyboarding was used to illustrate the environment and operations.

After performing a cognitive walkthrough, you'd typically want to perform more extensive tests by either adding more realism, or by adding more fidelity. More realistic tests are described in Section IV of this paper, where we present the current version of our core test battery, i.e. test battery AR (Abstract&Real-world). The tests in test battery AR must be performed in the real world and have low fidelity, meaning that they abstract from real world details.

Realistic tests can also be performed with higher fidelity, for example the NIST test arena [6]. NIST evaluates robots in test arenas that are as realistic as possible. Because the tests we propose in test battery AR are more abstract (i.e. have less fidelity) than the test in the NIST arena, they are a valuable contribution to NIST evaluation. Although it is very important to evaluate the robots also in a realistic environment, we state that for a good assessment, the environment and its resemblance are not always essential. To test the functions or requirements in an early phase design specification, it is only necessary to let the robot and the operator interact and operate on tasks that need similar capabilities as the real operational task [7]. This makes testing more easy (relatively cheap and under controlled conditions) to perform.

Another path to follow when testing UGV's in an iterative way is to perform tests in virtual environments, such as USARSim [3]. USARSim is a high fidelity virtual environment, used for research of human robot interaction in urban search and rescue. USARSim uses the Unreal Engine and accurately represents interface elements (such as camera) and robot behaviour, furthermore the users of USARSim can create their own robots and environments [8]. Eventually, robots should be evaluated in the real end-user environment, which is represented in the upper right corner of the graph.



### III. TEST DESIGN

This section describes how tests can be properly designed, regardless whether they are performed in the virtual world, or in the real world. We will first explain the general structure of the test battery. Then we will describe some general guidelines for test design, after which we discuss some issues regarding interpretation of UGV test performance.

#### A. Structure of the Test Battery

The tests in the test battery are categorized according to three levels in which robots are operated (from low-level operation to high-level operation):

- **Executorial:** At the execution level, the operator performs elementary actions, e.g., accelerating the robot, observing an object, etc.
- **Tactical:** At the tactical level, the operator executes a plan of actions. Most of the time, this is done during the mission. For example, the operator follows a route, defined using a number of waypoints.
- **Strategical:** At the strategic level, the operator forms a plan. For example, for the basic function transit this could be deciding on the waypoints for the route based on meteorological data.

Proper operator-robot interaction at these different levels requires different interface properties. For example, accelerating a robot (a task at the executorial level) requires the operator to be able to properly adjust the robot's speed. Deciding on the waypoints for a route (a strategical task), requires the operator to have spatial situation awareness of the area where the robot is to be operated. Therefore, the test battery should contain tests at each of the three levels.

Typically, higher level tests are dependent on the lower level tests. For example, the test "do a slalom" (tactical level) is dependent on the test "make a turn" (executorial level). If test A is dependent on test B, we mean that test A can only be passed if the agent also passes for test B. For each test in the battery, we the dependencies of the test must be made explicit.

#### B. Guidelines for Test Design

The experiment was within subject, and each participant first performed the test battery tasks, followed by the scenario.

#### C. Materials

A test can be viewed as a package containing a short name of the test, a unique identifier, a description of how the test is performed, the performance measures, the test's dependencies and a lab setup. Furthermore, we specify (in another table) the relation with operational demands. For the different elements of a test, related literature should be consulted on UGV metrics, testing environments, and operational literature.

Based on our experiences on designing tests for UGV's, we have identified the following aspects as deserving special attention:

#### *How is the test evaluated?*

In general, there are the following options:

- **Result-based:** Whether the operator is capable to accomplish a certain result. For example: can the operator do a slalom within 30 seconds? Within result-based evaluation we have two ways of doing that:
  - **Subjective:** an examiner determines whether the test has been passed
  - **Objective:** the test is passed by some objectively measurable criterium, such as time it takes to finish a trajectory.
- **Questionnaire-based:** useful for evaluating the operator's situation awareness. For example: ask the operator to draw a map of the environment after having moved around the environment for a while.

From our experiences with robot evaluation during the development of the test battery, we found that often a combination of result-based and questionnaire-based evaluation works well. For example, after the operator has finished a result-based test, we ask the operator what his or her experiences were and to estimate his or her performance on the test. It turned out that often the operator did not know that (s)he was disqualified for the test because (s)he was unaware of bumping into other objects during the task.

*Can the test be decomposed?* We aim for the tests to be as elementary as possible, following a reductionistic approach. This means that if we believe that a test actually measures two distinct independent aspects, we decompose the test in two separate tests for each of those two aspects. Of course, being able to do two things separately does not always imply that these two things can be done simultaneously as well. In those cases, we also include the test for doing the two things simultaneously in the test battery, provided that the composed task is realistic in an operational setting. The benefit of having the elementary tests as well is that it improves diagnostic power to our evaluation methodology.

*Is the test discriminatory?* Tests should be discriminatory in the following respects:

- It should not be a trivial test that every UGV can pass
- It should add something to the existing tests in the battery. The test should test a capability (or several capabilities) of the robot-operator system that in this combination are tested by no other test in the test battery.

#### D. Optimal Performance

To interpret the performance measure of a test, it is useful to understand what would be the *optimal* performance on the test. In general, the robot-operator performance is determined by two factors. Firstly, it is determined by the robot

properties. This is referred to as the *inherent capability*. For example, a robot can have an inherent capability of moving at a maximum speed of 20 kmh. Secondly, the robot-operator performance is determined by the operator controlling the robot. This is referred to as the *piloted capability*. The piloted capability can never be greater than the inherent capability of the robot. For example, an operator cannot make a robot go faster than its maximum speed. Typically, the piloted capability is lower than the inherent capability. For example, when doing a slalom, it is unlikely that the operator can move the robot at its maximum speed. In case the piloted capability equals the inherent capability of the robot, we can say that the test has been passed with *optimal performance*.

When the inherent capability of the different UGV's is different, the test results of the different UGV's cannot be straightforwardly compared. For example, steering an unmanned tank during a slalom task is much more difficult than steering a medium-sized robot, because it is a much larger robot. If the evaluator's interest is at the HRI properties of the two systems, this would be an unfair comparison. The comparison could be made more "fair" by comparing the actual performance of the test, with the optimal performance.

IV. CASE STUDY

We have applied our method in the domain of military UGV's. During the domain specification phase, we have identified the following four operational demands: *Transit*: The UGV should be able to transit from one location to another; *Observe*: contains all tests that are focused on the collection of mission relevant information; *Manipulate*: The manipulation of (objects in) environment, both direct (disposal of IED) and indirect (grenade throwing) manipulations are possible; *Communicate*: communication between operator/UGV with the other stakeholders.

Use cases developed for the NIFTI project, contain observe tasks. For instance when a UGV is deployed to retrieve the exact location of a victim in a collapsed building, the UGV will be tele-operated through an entrance (this can be a narrow passageway) and the UGV and operator will observe the environment for the victim by means of camera and sound.

A. Test Battery AR

The investigation of this case-study has resulted in a first version of a standard core test battery, i.e. test battery AR. All tests are abstract real world tests. A fragment of the core test battery is shown in Table IV-A.

For each of these tests, more details are provided on how the test should be carried out and which aspects are deemed important. For the slalom task, this is described in Figure 4.

B. Performing the tests

In our pilot experiment, we experimented with the Eye-robot (see Figure 6). The robot was operated from another

Nr	Name	Dependencies
<b>Transit Executional</b>		
TE1	Accelerate	-
TE2	Slow down	-
TE5	Accelerate backwards	-
<b>Transit Tactical</b>		
TT1	Do a Slalom	TE4, OE4, OT2
TT2	Stop before collision	TE1, TE2, OE1, OE4
TT9	Accelerate backwards following a straight line	TE5
<b>Transit Strategic</b>		
TS1	Find a way through the maze	TT1
TS3	Return to starting point	TE4, OS1
TS6	Estimate whether the UGV can drive up a slope	OE2

Table I  
FRAGMENT OF TEST BATTERY AR

<b>TT1</b>	<b>Do a slalom</b>
<i>Description</i> : Do a slalom between four traffic cones.	
<i>Lab-setup</i> : For the slalom test, we assume the following lab setup:	
<i>Performance measure</i> : time it takes to finish the slalom without hitting any of the traffic cones. In addition, the operator is asked the following questions: <ul style="list-style-type: none"> <li>- Did the UGV hit a traffic cone?</li> <li>- Did the UGV move a traffic cone (e.g. when turning around)</li> <li>- Did the UGV pass the cones narrowly or with a lot of distance?</li> </ul>	
<i>Motivation</i> : With this test, several capabilities are tested. First, it is tested whether the operator's situational awareness is good enough to do a <u>slalom</u> without hitting the cones. This includes controlling and adjusting the speed and the direction of the UGV according to the estimated position of the ...	
<i>Phase of Development</i> : Tested Concept	

Figure 4. Example of test description

room, using two possible user interfaces, i.e. a graphical user interface (GUI), and a telepresence interface (see Figure 5). The GUI interface consisted of a monitor with the robot's camera images and a joystick for steering. The telepresence interface is a more advanced interface, consisting of a head-mounted stereo vision device, together with a gas-pedal and a steering wheel for controlling the robot. Note that, in this paper, it is not our purpose to provide a complete investigation of these different interfaces. Rather, we wish to demonstrate our evaluation methodology by showing the feasibility of using early testing using the standardized test from test battery AR.

We have subjected both UI configurations to a selection of nine tests from the test battery. Some tests were discriminatory, among which TT1, TT2, and TS3. This means that they indicated a clear difference in performance between the GUI and the telepresence interface (mostly in favor of the



Figure 5. GUI interface (left) versus telepresence interface (right)

telepresence interface). Some of these differences appeared in the objective evaluation criteria specified with the test, such as the time it took to complete the slalom task. Other differences in performance were subjective, e.g., using the GUI interface the operator failed to notice that he bumped into one of the traffic cones. In conclusion, this case study of the methodology conveyed capability differences of two alternative user interfaces for a specific UGV with their operational impact.



Figure 6. The eye robot during the slalom test

## V. CONCLUSION AND FUTURE WORK

Standardized test methods for human-UGV interaction are important for Human Robot Interaction research. In this paper we have presented a usage-centered UGV evaluation method which is centered around a battery of basic tasks. We have also proposed guidelines to design appropriate tests, taking into account ease of testing, phase of UGV development, and the situatedness of the robot.

From our experiences with using our UGV test methodology, we believe that the benefits of using the method are twofold. Firstly, it provides a well-founded and structured way to perform a single UGV evaluation. It allows us to compare different UGV configurations (such as alternative user interfaces), and to benchmark designs. Secondly, it provides a good way to organize and store knowledge in our research team which is gained during prior UGV evaluations. For example, if we would experience that some maneuver is difficult for many UGV's, we would add a test to the standard battery that targets exactly this aspect. In this way, we can guarantee that this aspect is properly addressed in

next evaluations, and that it is recognized by all members of the UGV research team.

Addressing the interrelationships between the UGV, operator, interaction and context factors remains an important issue for further experimentation. In the future we plan therefore to specify and perform a summative experiment that is based on complex realistic use cases and utilizes the task battery knowledge-base. Also, we intend to extend the current test battery with tests that are to be performed in the virtual world. This would make our iterative testing method more complete, and would allow us to complement our existing abstract real world tests.

## ACKNOWLEDGMENT

This research is supported by FP7-EU grant for the NIFTi project (#247870FP7).

## REFERENCES

- [1] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: a survey," *Foundations and Trends in Human Computer Interaction*, vol. 1, no. 3, pp. 203–275, 2007.
- [2] T. Mioch, N. Smets, and M. Neerinx, "Predicting performance and situation awareness of robot operators in complex situations by unit task tests," in *Proceedings of the 5th International Conference on Advances in Human Computer Interaction (ACHI)*, 2012.
- [3] M. Lewis, "USARSim and HRI: from teleoperated cars to high fidelity teams," in *Proceedings of the IROS 2009 Workshop on Robots, Games, and Research: Success stories in USARSim*, 2009, pp. 18–22.
- [4] M. Neerinx, J. Lindenberg, N. Smets, A. Bos, L. Breebaart, T. Grant, A. Olmedo-Soler, U. Brauer, and M. Wolff, "The mission execution crew assistant: Improving human-machine team resilience for long duration missions," in *Proceedings of the 59th International Astronautical Congress (IAC2008)*, 2008.
- [5] N. Smets, J. Bradshaw, J. Diggelen van, C. Jonker, M. Neerinx, L. de Rijk, P. Senster, M. Sierhuis, and J. ten Thijs, "Assessing human-agent teams for future space missions," *IEEE Intelligent systems*, vol. 25, no. 5, pp. 46–53, 2010.
- [6] A. Jacoff, E. Messina, and J. Evans, "Experiences in Deploying Test Arenas for Autonomous Mobile Robots," in *Proceedings of the 2001 Performance Metrics for Intelligent Systems Workshop*, Mexico City, 2001.
- [7] J. W. Streefkerk, M. P. van Esch-Bussemaekers, M. A. Neerinx, and R. Looije, "Evaluating Context-Aware Mobile Interfaces for Professionals," in *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*. Information Science Reference, 2008, pp. 759–779.
- [8] J. Wang, M. Lewis, S. Hughes, M. Koes, and S. Carpin, "Validating USARsim for use in HRI Research," in *In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, 2005, pp. 457–461.

# An Error Analysis Model for Adaptive Deformation Simulation

Umut Koçak, Karljohan Lundin Palmerius, and Matthew Cooper

*C-Research*

*Linköping University, ITN*

*Norrköping, SWEDEN*

*Email: umut.kocak@liu.se, karljohan.lundin.palmerius@liu.se, matthew.cooper@liu.se*

**Abstract**—With the widespread use of deformation simulations in medical applications, the realism of the force feedback has become an important issue. In order to reach real-time performance with sufficient realism the approach of adaptivity, solution of different parts of the system with different resolutions and refresh rates, has been commonly deployed. The change in accuracy resulting from the use of adaptivity, however, has been paid scant attention in the deformation simulation field. Presentation of error metrics is rare, while more focus is given to the real-time stability. We propose an abstract pipeline to perform error analysis for different types of deformation techniques which can consider different simulation parameters. A case study is also performed using the pipeline, and the various uses of the error estimation are discussed.

**Keywords**—*physically based; deformation; multi-resolution; perception; error; analysis.*

## I. INTRODUCTION

The integration of haptics, the sense of touch, in human computer interaction has increased the immersion effect of virtual environments. Haptics has been used to serve different aims such as guidance, visualization, and realism in various computer applications. Considering the growing use of haptics in medical applications, such as in virtual simulations for training and rehearsal purposes, the degree of realism has become a crucial issue.

Achieving a compromise between realism and stability has always been a major challenge in haptics, especially for deformable objects. High refresh rates (1 kHz) are necessary to achieve a stable and continuous force feedback, while solution of the physical models with desired resolutions comes with a heavy computational burden. Adaptive multiresolution techniques have been among the most popular methods proposed to achieve desired refresh rates with required resolution. A major problem with this approach, however, has been a lack of focus on the reduced realism due to the error introduced by adaptivity, and the lack of standard error metrics.

In the literature it can be seen that consideration of the error is usually superseded by stability concerns but there are some error estimation techniques, widely studied in the context of Finite Element Methods (FEM) (e.g., [1], [2], [3], [4]), and adaptive studies [5] suitable for real-time use. Exploiting such methods to choose optimal solutions which

minimize error while maintaining sufficient performance to ensure stability would be beneficial. The common approach in the real-time deformation simulation field, however, is simply to include the maximum number of nodes which can be solved in real-time while keeping the simulation stable. Presentation of numerical values for the element sizes, time-steps and how they affect the accuracy is rare, however. One possible reason is the fact that estimating an error is not a trivial process, especially for adaptive simulations which vary parameters in real-time. In addition to deciding which parameter to consider (strain, stress, force) for error estimation, one also has to choose a reference. Therefore presentation of the numerical values without any error analysis is usually avoided in the literature.

In this study, we propose a pipeline used to analyze force feedback error caused by adaptive solution of deformation simulation. The idea depends on creating an offline error mapping of a deformation model for a set of parameters such as user inputs, material properties and adaptivity parameters. The potential uses of the offline error mapping include: (1) surveying correlations between error and parameters, (2) real-time quality assessment, and (3) real-time parameter adjustment to keep the error below some desired limits. The pipeline is applied in a case study by performing experiments in the simulation environment.

The outline of the paper is as follows: related work and theoretical background are presented in the second and third sections, respectively. The motivation for the work is discussed in the fourth section, followed by the explanation and the case study of the pipeline in the fifth and the sixth sections. The conclusions are presented in the final section of the paper.

## II. RELATED WORK

The main challenge in deformation simulation is reaching sufficient realism while maintaining the desired refresh rates. There are a variety of different deformation models and optimization techniques available. Among the deformation models, FEM is commonly considered to provide the most realistic behaviour and so, despite the substantial computational load it incurs, is the focus of most optimization and error analysis studies. In this section studies about adaptivity applied to FEM and the error analysis are summarized.

A common technique to achieve sufficient performance while retaining required realism around the point of contact is adaptivity. There are several approaches (for example [5], [6], [7]) which exploit adaptive spatial multiresolution models to keep the accuracy higher around the contact point while reducing computational load in the less precisely modelled parts.

Despite this frequent use of adaptive element size and time-step, the effects of such approaches on the perceived force feedback have not been comprehensively surveyed. The common method in real-time applications is, if error is considered at all, to use an error measure for each element and adapt the element size and the time-step for elements. The error estimation has, however, been more thoroughly studied in FEM (i.e. [1], [2], [3], [4]) although the real-time aspect is not, generally, as important there. Designing an ‘ideal-error-estimator-solution’ which is reliable and sufficiently computationally inexpensive that it can be used in real-time is an unsolved issue, as discussed in [3].

In this study, we propose a general pipeline to analyze the behaviour of the force error with respect to the level of detail in both time and spatial domains. The direct relationship between the force response and the solution of the whole model with different frequencies was presented in [8] for mass-spring models. In the pipeline we propose, the error can be analyzed for different deformation models and different parameters, such as material properties, and input types. An analysis of the error for a special case is also presented.

### III. BACKGROUND: STABILITY AND ACCURACY

The deformation simulation on a mesh depends on discretizing a continuous domain into elements. In addition, the iterative solution of the resulting differential equations depends on discrete time-steps. There are two main issues to be considered in the discretization: stability and accuracy. The adjustment of element size and time-step includes a compromise between these two concepts, therefore adaptive simulations should take these two concepts into consideration in real-time. Keeping the simulation stable means having a deformation model which always converges to the result. The accuracy, on the other hand, affects the error in the deformation.

#### A. Stability

The deployment of iterative solvers is generally preferred over analytical ones in deformation simulation because of their superior speed. The iteration process requires the choice of a time-step, which affects the accuracy and stability of the simulation. The two main categories of iterative solvers, implicit and explicit, have different behaviour in terms of stability. Implicit solvers are unconditionally stable, which means that the solution converges to the final value no

matter how large the time-step. On the other hand, the time-step is limited in the case of explicit solvers, where having an eigenvalue of greater than 1 in the stiffness matrix will result in divergence. The time-step limit depends on the maximum natural frequency which, in turn, depends on the material properties and the element size. The time-step limit can be physically interpreted to be that it must be small enough that no information propagates more than one mesh element per step. To achieve stability it is therefore necessary to consider the time-step and the element size together.

#### B. Accuracy

Increasing the time-step results in increasing error. In the case where a single ordinary differential equation is considered, the time-step is crucially important to achieve acceptable accuracy. In the case of a large number of differential equations, however, keeping the time-step just under the stability limit provides sufficient accuracy. The reason for this is the fact that a stiff system of differential equations covers a wide spectrum of natural frequencies. The stability limit is evaluated with respect to the highest frequency of the system. Therefore for a time-step which is close to the critical stability time-step limit, the response of the model for the highest frequencies will not have high accuracy. Fortunately, the structural response of the objects is dominated by much lower frequencies which are sufficiently more accurate for the chosen time-step. Therefore, for explicit solutions of a large number of equations, the accuracy is not an issue as long as the stability condition is satisfied [9].

The discretization in the spatial domain is another factor affecting accuracy. For instance, in case of FEM, the type, shape and size of the elements also play a role in the accuracy. Once the shape of the element is chosen, the order of the polynomial used to calculate the shape functions of the element can also vary. While increasing the order results in higher accuracy, using lower orders is common in real-time applications because of its drastic effect on the stability condition. The error of a quantity is of the order of  $\mathcal{O}(h^{p+1-m})$  [10]; where  $h$  is the element size,  $p$  is the degree of the polynomial used to calculate shape functions, and  $m$  is the order of the highest derivative in the governing equilibrium equation expressed in terms of displacements.

### IV. MOTIVATION

The common approach followed in deformation simulations is to choose the maximum time-step which guarantees stability for a given mesh resolution. As long as the stability conditions are satisfied, the accuracy criteria are also satisfied for large deformable meshes, as discussed in the previous section. It is not only the time-step but also the element size which affects the accuracy of the deformation behaviour. In real-time deformation simulation studies, it is usually considered to be sufficient to include the number

of nodes required for the mesh to be solved stably. The presentation of numerical values for the element sizes and how they affect the accuracy, however, is rare in this field.

Another issue which has been largely ignored in adaptive haptic applications where both the resolution and time-step are adjusted in real-time, is the potential to exploit the perception limits of human beings for optimization purposes. There have been numerous studies which have surveyed the different human perception limits but ways to exploit these limitations in hardware and software solutions are rarely explored. The research (for example [11], [12]) shows that there are just-noticeable-differences in force magnitude and in force direction. The threshold for force magnitude has been found to be around 7% in [11]. Similarly, the average force direction discrimination thresholds are 18, 26, and 32 degrees, depending on the accompanying visual input [12]. One could exploit these thresholds to save computational power by simplifying the force calculation while ensuring that the error in the force remains below the limit of perceivability.

In this work we propose a pipeline which analyzes the error due to the various parameters, including adaptivity parameters, and deploys a mechanism, without additional computational burden, to control the simulation parameters in real-time, according to the error criteria.

## V. PERCEPTUAL ADJUSTMENT MODEL

In order to introduce the facility to adjust the computational complexity without perturbing the user's perception of the represented forces it is necessary to examine the explicit correlation between force error and parameters, assessing quality of simulation at run-time, and adjusting adaptivity parameters to maintain the error under desired limits. The abstract pipeline is explained in this section while a case study performed by experiments is presented in the next section.

### A. Error Identification Principle

Achieving exact solution in real life scenarios is very rare for deformation simulations. There are several steps in the process contributing to the error, such as modelling, discretization and numerical errors. The pipeline proposed in this study focuses on the error caused by the deployment of adaptivity with varying element sizes and time steps in the simulation. In addition, a so-called goal-oriented [3] approach is followed by analysing the force feedback error of a contact node. In other words, the error considered in this model refers to the force error introduced by varying element sizes and time-steps while one must keep in mind the other contributions to the total error. For a given mesh and deformation model, the solution for the whole mesh with the highest resolution, and with a sufficiently low enough time-step to maintain stability, is considered as a 'reference' force with no error.

### B. Offline Error Characterization

The idea of calculating an offline error mapping depends upon applying Monte Carlo simulation for a set of parameters to obtain a force feedback value for each sample. One needs to maintain stability while varying the tested parameters in a reasonably wide range. To maintain stability, the time-step needs to be smaller than a critical time-step which depends on a number of parameters including the element size, elasticity, mass and poisson ratio. This causes different stability behaviour in the simulation for different materials. The choice of time-step for real-time solutions is also dependent on the computational power available and the number of nodes required. Considering the limitations on the time-step and its effects on the range of parameters to be tested, offline deformation solutions are preferred to create the error mapping. This allows the exploration of time-steps decoupled from the limits imposed on computational power and number of nodes in addition to maintaining the stability for a wide range of parameters for a given time-step.

The idea of offline error mapping is illustrated in Figure 1. The pipeline uses a number of types of input parameters to a chosen deformation model and creates an error mapping specific to these parameters which can be categorized as user input, material properties and adaptive simulation parameters. The amount of strain, input frequency, and choice of contact node can be named as examples of user input. The material properties include the physical properties of the object such as elasticity, mass, poisson ratio, or damping, while the deformation model can be any deformation algorithm such as FEM, mass-spring etc. The adaptive simulation parameters include the distribution of element sizes and time steps throughout the mesh.

The first step in creating the error mapping is an offline evaluation of the reference force explained in Section V-A for a range of user input and material properties parameters. The force feedback is then calculated again for the same set of parameters but deploying adaptivity with varying element sizes and time-steps. The force feedback obtained by deploying adaptivity for this set of parameters is compared with the corresponding reference force, evaluated without adaptivity, to calculate the error. This procedure applied to a specific case is explained in more detail in Section VI-C.

### C. The Uses of Error Mapping

One of the potential uses of an error mapping is surveying explicit correlation between force error caused by adaptivity and material or input parameters. This might be used to gain more insight about the effects of adaptivity for different deformation models and materials leading to better choices of algorithms for different types of applications. Another use would be assessing quality of simulation at run-time without a major computational burden. Since the error mapping has been pre-calculated offline, the only extra work introduced is to find the corresponding error in real-time, by using a

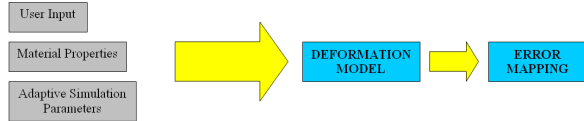


Figure 1. The offline error mapping procedure: The user input, material properties and adaptive simulation parameters are given as input to the deformation model. The force feedback for each trial is compared with the solution of the whole object with highest resolution and an error mapping is created for each trial.

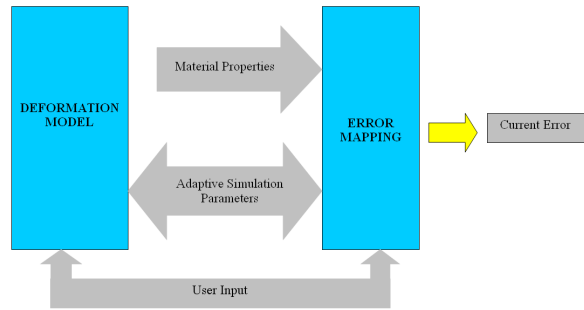


Figure 2. The online error assessment and quality adjustment: For a given set of material properties, adaptive simulation parameters and user input, the error mapping is used to determine the current error. This can either be used purely for quality assessment purposes or the adaptive simulation parameters can be fed back to the deformation model at run-time aiming for a compromise between target error value and stability.

lookup table, as described in the results section. Figure 2 illustrates the online uses of the error mapping. In addition to just evaluating the current error for assessment purposes, the adaptive simulation parameters can be adjusted and fed back to the deformation model in order to maintain the error under desired limits. This real-time adjustment, however, has to consider the stability issues which are coupled with the available computational power to achieve a compromise between error and stability. In the event that both constraints cannot be simultaneously met during real-time use, the system must ensure that stability is retained while recording the fact that the error limitations have been exceeded and, perhaps, informing the user that this failure has occurred.

## VI. EXPERIMENTS

Experiments have been performed to apply the abstract pipeline proposed to a specific scenario for a chosen deformation model and a set of parameters. We foresee no obstacles to applying the same principles and performing a more elaborate analysis by extending the number of parameters.

Linear FEM, where performance is a less serious concern compared with non-linear FEM, was chosen as the deformation model. The adaptive solutions, however, are still needed for larger numbers of nodes even for linear FEM. Considering the deformation characteristics in real life situations, complex properties such as viscoelasticity

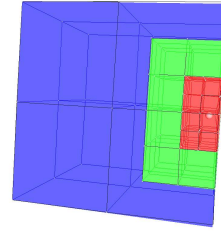


Figure 3. The allocation of asynchronous regions in the object. Different regions have different mesh resolutions in addition to being solved with different time-steps. The primary region (local neighbourhood of the contact) is shown with red, and the secondary regions are shown in green and blue.

and anisotropy are also commonly observed in addition to non-linearity. The error analysis for tissue deformation is, therefore, a non-trivial process which needs to consider different physical properties. The reason we chose the linear FEM is to make the proof of concept of the pipeline in a relatively simpler model, and analyze more complex models as future work. Clearly the structure of the pipeline presents no obstacles to either changing the deformation model or extending the set of parameters.

### A. Deformation Model

A dynamic linear finite element model with a cubic element is used. The simulation time is discretized into time-steps to evaluate the numerical integration of (1):

$$\mathbf{M}\ddot{u} + \mathbf{C}\dot{u} + \mathbf{K}u = f \quad (1)$$

where the stiffness, mass and damping matrices are represented by  $\mathbf{K}$ ,  $\mathbf{M}$ , and  $\mathbf{C}$  respectively, while  $u$ ,  $\dot{u}$ ,  $\ddot{u}$  and  $f$  refer to displacement, velocity, acceleration and force vectors. The mass matrix is diagonalized and a constant damping for each node is used. The reader is encouraged to see [10] for further details of the finite element method.

### B. Parameters

**Adaptive Simulation Parameters:** There are different ways to implement adaptive deformation (for example [5], [13]). In our framework the deformation is calculated with different resolutions and at different update rates for different regions of the object. The local neighbourhood of the contact point maintains higher resolution and smaller time-steps than the more remote regions, illustrated in Figure 3. The corresponding FEM equations for different regions are solved asynchronously with different time-steps and resolutions. The regions, therefore, are referred to as *asynchronous regions*. Throughout the rest of the paper the local neighbourhood of the contact node, having the highest frequency, will be referred as *the primary region* and the others as *the secondary regions*. The framework used allows real-time adaptation of regions and their sizes

as the nature of the contact is changed. The details of the implementation, however, are out of the scope of this study and the reader can see different examples of adaptive simulations (for example [5], [13]) for further details.

In this system, the force at the contact node depends on the resolutions, time steps and the sizes of the regions. Two asynchronous regions have been employed throughout the experiments presented here. The primary region's spatial and time resolution has been fixed for all experiment sets. The spatial resolution (2.5 cm) has been chosen to be the highest resolution used during the reference force calculation without adaptivity and the time-step has been kept at 1 kHz throughout for the primary region.

The secondary region frequency is one of the parameters being varied using rates of 1 kHz, 900, 800, 700, and 600 Hz. The use of lower time-steps was possible for some of the data sets, however not for all. For a given resolution, for instance, the refresh rates need to be higher for more stiff materials, or lower mass densities. In order to be able to analyze a wider range of parameters, higher stiffness or lower mass densities, with the same frequency values the refresh rates were chosen to be high. As explained in Section III-B, the accuracy is not affected by the time-step as long as the stability criteria are met. This is also observed in the results showing that the frequency does not affect the error. The choice of high refresh rates to cover a wider range of material properties therefore has no effect on the error analysis.

The resolution of the the secondary region is another parameter under consideration and has been varied over multiples of the primary region resolution which are  $1\times$ ,  $2\times$ , and  $3\times 2.5$  cm.

The third parameter surveyed is the size of the primary region. The size has been measured as the number of nodes in all directions from the contact node homogeneously and changed between 1, 2, 3 and 4 in three dimensions. To illustrate, a region with value of 1 as size includes  $3\times 3\times 3$  nodes in three dimensions, including the contact node.

#### User Input:

To apply a deformation to the model, a simulated haptic device was used allowing the application of a precise and reproducible input. While it is possible to apply deformation to a chosen set of contact nodes, the deformation was applied to a fixed contact node throughout the experiments. Two parameters have been changed for the deformation applied: Strain and frequency. To obtain a realistic response from linear FEM the maximum amount of strain applied has been kept limited, typically to less than 10% [14] of the mesh size. The amount of strain has been chosen to be 0.5, 1.0 or 1.5 cm. The frequency of the input has been changed between 0.5, 1.0 and 1.25 Hz.

**Material Properties:** Physical properties like elasticity, mass, damping and poisson ratio have been varied during the experiments. The ranges have been chosen according

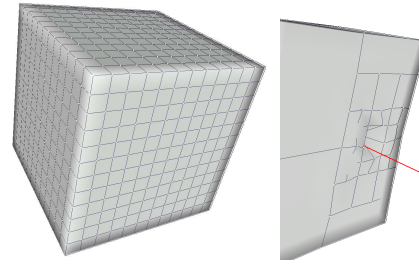


Figure 4. The experiments have been performed with a cube composed of  $13\times 13\times 13$  nodes. The deformation is shown on the right half side of the figure.

to the soft tissue properties described in the literature. The elasticity values measured in the different studies are: between 5 and 35 kPa in [15], 6 and 11 kPa in [16], 0 and 3.5 kPa in [17]. The differences between the results of the studies mentioned are due to the tissue type, the magnitude of deformation, the non-linear behaviour of the elasticity and the level of precompression as discussed in [18]. Consequently, the elasticity has been set to 5, 10, 15 and 20 kPa in the experiments. The mass density has been found to be  $1\text{ g/cm}^3$  in [19] and was set to 0.8, 1.0 and  $1.2\text{ g/cm}^3$  for the experiments. Measurement and modelling of the damping for soft tissue is non-trivial since the coefficient depends on stress level [9] and vibration frequency. The damping for a human thigh has been measured for different applied force magnitudes and frequencies and angles of flexion of the knee in [20]. The damping varies between 7 and  $102\text{ s}^{-1}$  for a force applied normal to the skin. The response of the deformation behaviour in pilot studies was also considered while deciding the damping values and they were set to 10, 20, 30 and  $40\text{ s}^{-1}$  during the experiments. Poisson ratios of both fat and muscle tissue have been measured to be approximately 0.49 in [21] and have been set to 0.40, 0.45 and 0.49 for the experiments.

#### C. Procedure

A deformation is applied to the model by a simulated haptic device which provides a precise and reproducible input. A reference force is evaluated at the contact node by solving the whole mesh without any adaptivity. The same input is then applied by employing two asynchronous regions with varying parameters. The force responses of the asynchronous regions are compared to the reference force to obtain an error value. This procedure is repeated for the full set of ranges of material user input and material parameters.

A cube, shown in Figure 4, with edge length of 25 cm and  $13\times 13\times 13$  nodes is used for the experiments. The four corner nodes on the back face are fixed and input is applied to the middle node on the front face. For deformation, an explicit central-difference solver is employed with cubic finite element type together with linear FEM.



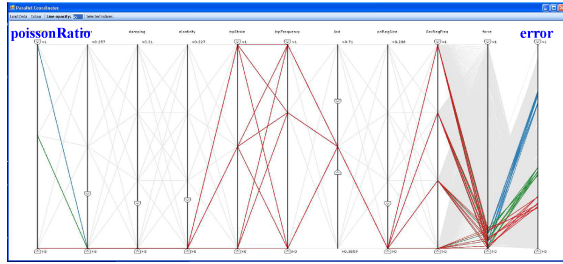


Figure 5. Parallel coordinates allow restriction of some parameters in order to observe the effects of the others. The use of colours enables the observation of how each parameter affects the error. In this case the material properties other than poisson ratio, resolution of secondary region and primary region size are fixed, and colours are used to distinguish the effect of poisson ratio on the error. The right-most axis is error and the left-most the poisson ratio. The colours identify three different levels of poisson ratio.

**D. Results**

The error mapping obtained can be used to analyze the effect of each parameter on the force, separately. The correlations between the error and the individual parameters are presented in this subsection. A data set consisting of 62208 sample points has been obtained by a Monte Carlo simulation. After elimination of the divergent sets, the number was reduced to 54083 samples. Visualization techniques and statistical tests have been used to analyze the data.

Parallel coordinates, as illustrated in Fig. 5, are used to visualize the effect of each variable on the force. The use of parallel coordinates allows the user to keep all of the parameters but one fixed or restricted so that the effects of the free parameter on the error can be observed. Parallel coordinates provide a general picture of how the error is affected by each parameter. However, to convey a quantified measurement between the error and the parameters, statistical tests can be performed. Two different partial correlation tests have been carried out.

To observe how the error is affected by a single parameter, eliminating the effects of the other parameters, two common partial correlation tests are applied. Table I illustrates the correlation coefficients of error and the p-values for each parameter obtained by the Pearson and Spearman tests. The former is suitable for linear relations and assumes normal gaussian distribution for each variable. Since the ranges used during the experiments do not meet the gaussian distribution requirement, and the linearity of the relations cannot be known in advance, the Spearman test is more suitable for the data. If the Spearman parameter is larger than the Pearson in absolute value, however, one can conclude that the variables are consistently correlated but not in a linear fashion. The correlation coefficient varies between -1 and 1. The higher the absolute value of the coefficient the more correlated the parameters are with the error. To determine the significance of covariance, a t-test with a confidence level of 95% is typically applied, the p-values of which are also presented in

Table I  
THE PARTIAL CORRELATION BETWEEN ERROR AND THE PARAMETERS

	Error Correlation			
	Pearson		Spearman	
	Coefficient	p	Coefficient	p
Poisson Ratio	0.5837	0	0.5488	0
Mass	0.0166	0.0001	0.0053	0.2152
Damping	0.0216	0	-0.1691	0
Elasticity	0.4746	0	0.2712	0
Input Strain	0.4377	0	0.158	0
Input Frequency	0.0119	0.0058	-0.1244	0
Level of Detail	0.4582	0	0.7033	0
Pr. Reg. Size	-0.5624	0	-0.6243	0
Sec. Reg. Freq.	-0.0015	0.7354	0.0003	0.9403
Force	-0.5073	0	-0.1827	0

Table I. P-values below 0.05 can be interpreted as indicating that the correlation is very unlikely to be different from zero by chance. The correlation coefficients are interpreted as small between 0.0 and 0.1, medium between 0.1 and 0.3, and large between 0.5 and 1.0 [22].

For the scenario considered in these experiments, the analysis of the correlation indicates the following. Amongst the material properties, the poisson ratio is the one most correlated with the error with a p-value of < 0.01. Increasing the poisson ratio causes the error to increase significantly. One cannot say the same for the mass from the p-value. Therefore parallel coordinates are used to manually check the relationship between the mass and the error, and it can be said that no visually observable significant change occurs in error due to changing mass. The elasticity has a positive, and the damping has a negative correlation with the error, but neither is as strong as that of the poisson ratio.

Of the input parameters, frequency and strain show negative and positive correlation, respectively. The strength of the correlations are not, however, significant and they can be grouped into medium correlation.

One can see the high positive and negative correlation with the element size of the secondary region and the primary region size, respectively. Both correlations can be said to be nonlinear since the Spearman coefficient is larger than the Pearson. The refresh rate of the secondary region has no correlation with the error. This can be explained by the fact that, as explained in Section III-B, a small enough time-step for stability also satisfies the accuracy criteria as long as the solution system is stiff enough [9].

**E. The Real-Time Uses of Error Mapping**

The error mapping can be used to determine the force error in real-time. This can allow some simulation parameters to be adjusted to keep the error below limits such as those set by human perception. The adaptive simulation parameters, such as element size and time-steps, can be

adjusted depending on the error. The material properties are measured physical values and the input parameters are determined by the user. These two should be used in obtaining the real-time error from the mapping rather than being adjusted to achieve a desired error.

Real-time determination of the error from the error mapping might be non-trivial and depends on how the mapping is obtained. A look-up table, as shown in Table II, is then used to determine the error for the current parameters and simulation state. Some issues such as continuously varying input frequency, heterogeneous material properties or changing contact node might require special solutions during the error mapping lookup phase.

The real-time measurement of input frequency may not be practical, instead one can exploit the fact that human beings have limited bandwidth while touching objects. An upper bound for a force control was found to be 25 Hz in [23] and 20 Hz in [24]. From the human output perspective 10 Hz is considered to be more than sufficient according to [25]. For the presented error mapping in the example, the lowest input frequency results in the highest error. Therefore the lowest frequency can be considered to determine a maximum bound for the effect of input frequency on the error.

Another issue is the deformation of objects with heterogeneous material properties making the use of a look-up table non-trivial. One alternative to address this problem is to check the material properties within a local neighbourhood and consider the combination of material properties with the highest error in the neighbourhood. For example considering the maximum elasticity, poisson ratio and minimum damping will ensure the error constraint is on the safe side by assuming the worst case scenario. Another alternative, which is less safe, would be to take average of each material property within the local neighbourhood to determine error. The limitation of this technique is the existence of rigid sub-objects, the effects of which on asynchronous regions will be explored in future studies.

The number of contacts considered during the creation of the error mapping determines how elaborate the mapping is. Creating an error mapping for each node in the mesh can be quite expensive. A mapping with few nodes and real-time interpolation of the error between these nodes provides an error value for each node of the mesh.

## VII. CONCLUSIONS AND FUTURE WORK

In this study, we have proposed a pipeline to survey the varying accuracy of the force feedback due to adaptivity with respect to a number of parameters of a deformation model. First, an error mapping with respect to certain number of parameters, including user inputs, material properties and adaptive simulation parameters is created. The deformation is first solved off-line for the whole object with the highest resolution mesh and the force response saved as a reference. The adaptivity is then employed with different parameters

Table II  
THE LOOK-UP TABLE FOR THE ERROR

Material Prop.				Applied Input		Asynch. Reg.		
P.R.	M	D	E	Strain	Freq.	Lod	Pr.Reg	Err
0.49	0.8	10	5000	0.005	0.5	2	1	56.7
0.49	0.8	10	5000	0.005	0.5	2	2	15.7
0.49	0.8	10	5000	0.005	0.5	2	3	8
0.49	0.8	10	5000	0.005	0.5	2	4	1.6
0.49	0.8	10	5000	0.005	0.5	3	1	66.7
0.49	0.8	10	5000	0.005	0.5	3	2	60.4
0.49	0.8	10	5000	0.005	0.5	3	3	6.8
0.49	0.8	10	5000	0.005	0.5	3	4	4.1
0.49	0.8	10	5000	0.005	1.0	2	1	60

and the force output is compared with the reference force to estimate an error for the given parameters. This procedure is repeated for a range of different material properties, and user inputs. In addition to providing an error measure without any serious real-time computational burden, the offline solution of the deformation decouples computational power from the stability requirement. The whole stable range of parameters for a given deformation model can, therefore, be solved offline and conveyed through the error mapping.

The error mapping obtained offline can be used to survey the correlation between force error and material properties, user input and adaptive simulation parameters. One can compare how different deformation models, such as linear FEM, non-linear FEM, or mass-spring are affected by adaptivity. This exploration has the potential to help in choosing algorithms and adaptivity parameters for different types of application. At run-time the quality of the force feedback can be continuously monitored with the help of error mapping, allowing the adaptive parameters to be adjusted to keep the error below the desired limits while maintaining stability.

Experiments have been performed on an example application of the proposed pipeline. The error mapping was created for a linear FEM. Material properties such as elasticity, mass density, damping, poisson ratio and user input parameters like strain, input frequency have been changed in addition to adaptive simulation properties. Observed correlations between the error and the input parameters have been presented and an example use of a look-up table for error mapping in real-time was proposed.

One must keep in mind that the example presented in the experiments section is a special case of the pipeline proposed in this study. In real life situations, more complex material properties such as nonlinearity, viscoelasticity, or anisotropy are commonly observed. Consequently achieving an ideal error estimation is non-trivial [3] in deformation simulation. We used a linear FEM to visualize an application of the pipeline because of its relatively simple structure. The possible challenges that might occur during the real-time use of the error mappings for models exhibiting material inhomogeneity, interpolation of error for contact nodes, and corresponding solutions have also been discussed.

In future work we will study the pipeline with alternative parameters and deformation models. The effect of different contact nodes, shapes, the use of mapping for heterogeneous materials, nonlinearity, and viscoelasticity can be named among the concepts to be surveyed. A user-centred study employing the error mapping with just noticeable differences will be very interesting to study in the subsequent phases.

#### ACKNOWLEDGMENTS

We thank Camilla Forsell for her help with the statistical analysis applied in this work. This work has been funded by the Swedish Science Council through grant number 621-2005-3609, the Foundation for Strategic Research (SSF) under the Strategic Research Center MOVIII, and the Swedish Research Council Linnaeus Center CADICS.

#### REFERENCES

- [1] M. Bercovier and O. Pironneau, "Error estimates for finite element method solution of the stokes problem in the primitive variables," *Numerische Mathematik*, vol. 33, pp. 211–224, 1979.
- [2] E. Rank and O.C. Zienkiewicz, "A simple error estimator in the finite element method," *Communications in Applied Numerical Methods*, vol. 3, pp. 243–249, 1987.
- [3] T. Gratsch and K.J. Bathe, "A posteriori error estimation techniques in practical finite element analysis," *Computers & Structures*, vol. 83, pp. 235–265, 2005.
- [4] M.J. Aftosmis and M.J. Berger, "Multilevel error estimation and adaptive h-refinement for cartesian meshes with embedded boundaries," in *40th AIAA Aerospace Sciences Meeting and Exhibit, AIAA Paper*, 2002.
- [5] G. Debonne, M. Desbrun, M. Cani, and A. Barr, "Dynamic real-time deformations using space & time adaptive sampling," in *Proc. Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH 01)*, 2001, pp. 31–36.
- [6] S. Capell, S. Green, B. Curless, T. Duchamp, and Z. Popović, "A multiresolution framework for dynamic deformations," in *Proc. SIGGRAPH/Eurographics Symposium on Computer Animation*, 2002, pp. 41–47.
- [7] E. Grinspun, P. Krysl, and P. Schroder, "Charms: A simple framework for adaptive simulation," in *Proc. Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH 02)*, 2002, pp. 281–290.
- [8] M. Cavusoglu and F. Tendick, "Multirate simulation for high fidelity haptic interaction with deformable objects in virtual environments," in *Proc. Robotics and Automation*, 2000, pp. 2458–2465.
- [9] R.D. Cook, D.S. Malkus, and M.E. Plesha, *Concepts and Applications of Finite Element Analysis*, 3rd ed. John Wiley and Sons, 1989.
- [10] O.C. Zienkiewicz, R.L. Taylor, and J.Z. Zhu, *The Finite Element Method: Its Basis and Fundamentals*, 6th ed. Elsevier Butterworth Heinemann, 2005.
- [11] X.D. Pang, H.Z. Tan, and N.I. Durlach, "Manual discrimination of force using active finger motion," *Perception and Psychophysics*, pp. 531–540, 1991.
- [12] F. Barbagli, K. Salisbury, C. Ho, C. Spence, and H.Z. Tan, "Haptic discrimination of force direction and the influence of visual information," *ACM Trans. Appl. Perception*, vol. 3, pp. 125–135, 2006.
- [13] U. Koçak, K. Palmerius, and M. Cooper, "Dynamic deformation using adaptable, linked asynchronous fem regions," in *Proc. Spring Conference on Computer Graphics*, 2009, pp. 213–220.
- [14] G. Picinbono, H. Delingette, and N. Ayache, "Non-linear anisotropic elasticity for real-time surgery simulation," *Graph. Models*, pp. 305–321, 2003.
- [15] V. Egorov, S. Tsyuryupa, S. Kanilo, M. Kogit, and A. Sarvazyan, "Soft tissue elastometer," *Medical Engineering and Physics*, pp. 206–212, 2008.
- [16] X. Li, G. Wang, L. Huang, and G. Zhang, "Young's modulus extraction methods for soft tissue from ultrasound measurement system," *Instrumentation Science and Technology*, pp. 393–404, 2006.
- [17] R. Sinkus, J. Lorenzen, D. Schrader, M. Lorenzen, M. Dargatz, and D. Holz, "High-resolution tensor mr elastography for breast tumour detection," *Physics in Medicine and Biology*, pp. 1649–1664, 2000.
- [18] A. Samani, J. Bishop, C. Luginbuhl, and D.B. Plewes, "Measuring the elastic modulus of ex vivo small tissue samples," *Physics in Medicine and Biology*, pp. 2183–2198, 2003.
- [19] P.C. Johns and M.J. Yaffe, "X-ray characterisation of normal and neoplastic breast tissues," *Physics in Medicine Biology*, pp. 675–695, 1987.
- [20] J.M. Wakeling and B.M. Nigg, "Soft-tissue vibrations in the quadriceps measured with skin mounted transducers," *Journal of Biomechanics*, vol. 34, pp. 539–543, 2001.
- [21] G. Rus and J.G. Martinez, "Ultrasonic tissue characterization for monitoring nanostructured tio2-induced bone growth," *Physics in Medicine and Biology*, vol. 52, pp. 3531–3547, 2007.
- [22] A. Field, *Discovering Statistics Using SPSS*, 2nd ed. Sage Publications Ltd, 2005.
- [23] R.N. Stiles and J.E. Randall, "Mechanical factors in human tremor frequency," *Applied Physiology*, vol. 23, pp. 324–330, 1967.
- [24] M.A. Srinivasan and J.S. Chen, "Human performance in controlling normal forces of contact with rigid objects," in *Winter Annual Meeting of the American Society of Mechanical Engineers*, 1993, pp. 119–125.
- [25] T.L. Brooks, "Telerobotic response requirements," in *IEEE Int. Conf. on Systems, Man, and Cybernetics*, 1990, pp. 113–120.

## On the Evaluation of Auditory and Head-up Displays While Driving

Christina Dicke  
 relevantive AG  
 Berlin, Germany  
 christina.dicke@relevantive.de

Grega Jakus, Sašo Tomažič and Jaka Sodnik  
 Faculty of Electrical Engineering  
 University of Ljubljana, Slovenia  
 grega.jakus@fe.uni-lj.si  
 saso.tomazic@fe.uni-lj.si  
 jaka.sodnik@fe.uni-lj.si

**Abstract** — In this paper, we propose a low cost, laboratory based testing framework for in-vehicle interfaces. Exemplified by a comparison between an auditory interface, a Head-up display, and a combination of both we show how task completion times, driving penalty points, mental workload, and subjective user evaluations of the interfaces can be collected through different logging systems and user questionnaires. The driving simulator used in the experiment enables the simulation of varying traffic conditions as well as different driving scenarios including a highway and a busy city center. Only some preliminary results are reported in this paper.

**Keywords**-Human-computer interaction; auditory interface; head-up display; car simulator; driving performance.

### I. INTRODUCTION

Driver distraction is estimated to contribute to up to 25 percent of vehicle crashes [1]. It is defined as the direction of attention away from safely handling the vehicle towards an object or event in the internal or external vehicle environment [2]. In many cases this distraction originates from in-vehicle tasks that are unrelated to driving such as making a phone call, sending text messages, or adjusting controls within the car. Distraction generated by interacting with smartphones is likely to increase as these devices are not only becoming more popular and affordable, but also offer services tailored towards in-vehicle use such as GPS (Global Positioning System) based navigation and real time traffic information [3].

According to the multiple resource theory of attention [4], humans only have limited amounts of attention available at any given time. Different tasks can use different attention resources or share them. If resources are shared interferences may occur leading to decreased performance in all tasks. For example, driving a car demands a significant amount of visual attention. Operating a navigation system or mobile phone through a visual interface while driving competes for the same resource associated with visual perception and is therefore likely to cause distraction from the primary (driving) task [5] [6]. It has also been shown that physical and cognitive distraction significantly impair the driver's visual search patterns, reaction times, decision-making processes, and the ability to maintain speed, throttle control, and lateral position on the road [1][7].

Integrating the smartphone handset with in-car electronics and merging the access to all car and mobile device functions in one haptic interface, such as a multifunctional steering wheel, can help to reduce the haptic distraction. Head-up displays (HUD) have been proposed as a solution to reduce the frequency and duration of the driver's eyes-off-the-road by projecting information on the windshield. HUDs, when

compared to HDD (Head-Down Displays), have been shown to reduce the response times to unanticipated road events and lead to smaller variances in lateral acceleration and steering wheel angle [8]. However, they have also been shown to increase mental load as indicated by longer response times in high workload situations [9][10].

To reduce visual distraction, speech-based interface have been proposed [11][12] as they demand resources associated with auditory perception and are therefore less detrimental [13]. However, extensive user testing with particular emphasis on the evaluation of the distractive potential and the user experience of individual services or interfaces is crucial and can only be partially derived from prior research. Therefore, we propose a testing framework that allows for low cost, laboratory based usability testing. Using the example of a comparison between three different interfaces for interacting with an in-vehicle communication, navigation, and entertainment system while driving, a HUD display, an auditory interface, and a combination of both, we demonstrate how these interfaces can be prototypically realized and thoroughly evaluated.

### II. USER STUDY

The design rationale of this study is to investigate the impact of multimodal interfaces for in-vehicle control systems on driving and task solving performance. In particular, an audio only interface is compared to a visual only HUD and a combination of both. To simulate a realistic use case, the experiment is running in a driving simulator. Participants perform tasks of different complexity while they drive the simulated vehicle on different routes and with different traffic conditions. The driving simulator depicted in Figure 1 consists of a large projection screen, a steering wheel, accelerator and brake pedals, and the HUD shown in Figure 2.



Figure 1. The car simulator consisting of a large projection screen, a steering wheel, foot pedals and a gear stick.

By means of a custom-made interaction device attached to the steering wheel (shown in Figure 3, left) participants can navigate through a menu structure as illustrated in Figure 3, right.



Figure 2. The visual interface implemented as a HUD.

Three different experimental conditions are created by using three different user interfaces. The same menu structure is used with all three interfaces. There are up to 8 items on each level with the top level containing the following items:

- Heating and Cooling
- Entertainment
- Communications
- Navigation and Traffic
- Trip computer

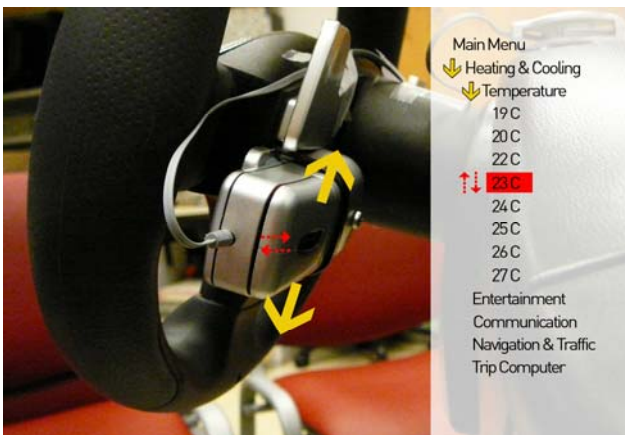


Figure 3. The interaction device and its use for navigating the menu structure.

### III. EXPERIMENT DESIGN

The study design is a within subjects 3 x 2 (Interface x Task Complexity) design with participants randomly allocated to one of six groups. To prevent learning effects each group has a different combination of route difficulty and conditions.

For example, one group starts with the auditory interface in a low traffic situation, proceeds with the visual interface in a low traffic situation, and ends with the audio-visual interface in a high traffic situation. A second group starts with the

audio-visual interface in a low traffic situation, proceeds with the visual interface in a high traffic situation, and ends with the auditory interface in a high traffic situation and so on.

#### A. Experiment Apparatus

1) Computer Setup: Three different computers are used for the experiment running:

- A driving simulator (DS),
- User interaction (UI) application and
- Management and logging (ML) software suite

The computers are communicating via TCP/IP protocol stack in a private network. The DS machine is used for running the driving simulator and for logging driving errors.

The UI machine is running the user interaction application and constantly reports the events connected with the interaction device (button clicks, mouse wheel turns) to the ML computer.

The ML computer is used as a main machine for conducting the experiment. The operator of the ML machine can control the driving simulator using remote desktop software. The ML computer is also used for collecting driving and interaction device events, measuring task completion times, and for filling-in the user questionnaires.

The software running on the UI and ML machines was developed on the Java platform. The input data required by the UI application are stored in an XML (eXtensible Markup Language) file. The file contains hierarchical menu structure, textual content that is displayed in visual interface, and references to pre-recorded sound clips for the auditory interface.

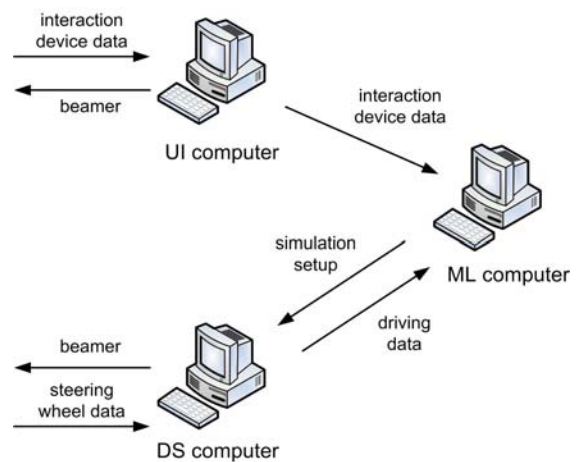


Figure 4. Experiment setup

2) Beamers: The visual interface and the image of the driving simulator are projected on a projection screen using two beamers. The beamer projecting the simulator image is mounted on the ceiling of the room, while the beamer projecting HUD is placed on the driver's right side in the height of his/her shoulders.

3) Car Simulator: The essential component of the experiment is a driving simulator software. We chose City Car Driving version 1.2 [14], which was projected on the big screen. It

supports the simulation of multiple driving environments, such as different regions of a city center, a motorway or a highway. It also enables a variety of different driving routes and traffic intensity conditions. The Thrustmaster’s “RGT Force Feedback CLUTCH Racing Wheel” is used as the input device, which comprises of a steering wheel, three foot pedals and a gear stick. Car and environment sounds are played using Genelec’s “8030A bi-amplified monitoring system” consisting of two loudspeakers placed on both sides of the projected image.

Test subjects are driving a left-handed Peugeot 206 CC with automatic gearbox. The traffic is right-handed. The route named “Motorway” with 10 percent traffic intensity is used for the “low traffic” condition while the route named “Modern district” with 50 percent traffic is used for “high traffic” condition. Test subjects are given loose navigation instructions ensuring comparable driving experience.

**B. Visual Interface**

The visual interface (Figure 2) is a 20 x 20 cm HUD projected to the right-central position of the windshield (above the car’s central LCD display which was not used in the study). The HUD displays five of the available items of the selected menu or submenu. When there are more than five items available, the user can access them by “scrolling” up or down in the current menu level. The menu is designed to be non-circular with the peg at both ends of the menu or submenu. The menu items are displayed in a high contrast yellow color. The selected item is highlighted with red fonts with slightly increased font size compared to non-selected items. On the top of the HUD a green colored title indicates the currently selected submenu. The setup was designed and programmed to allow for a quick adjustment and accommodate required changes of the menu structure, position, size, and content.

**C. Auditory Interface**

The auditory interface is based on prerecorded sound clips generated by AT&T Labs TTS Demo (text-to-speech) technology. A male voice called “Mike” is used for the main menu structure while other voices are used for imitating various tasks (voicemail messages, traffic report service, etc.). The OpenAL and JOAL libraries are used for the creation of dynamic sound sources. Sounds are played through two computer loudspeakers, which are placed at the usual position of car speakers mounted in vehicle doors.

**D. Interaction Device**

The navigation within the menu is enabled by using a custom-made interaction device – a small mouse attached to the steering wheel (Figure 3, left). The interaction device consists of two buttons and a scrolling wheel, which can also act as a third button if pressed. The scrolling wheel is used to navigate among the items available at a certain level of the menu structure. If the scrolling wheel is pressed while using the acoustic interface, the title of the current submenu is played. The other two buttons are used to confirm the current selection or to exit the current submenu and move up one level in the menu structure (cf. illustration in Figure 3).

**E. Experiment Conditions and Tasks**

Three different experiment conditions were defined. The conditions “A” and “V” are based on the acoustic and visual interfaces described in previous sections while the condition “AV” is based on the combination of both. Five tasks are performed within each experimental condition. Each group of tasks consists of three simple “atomic” and two “complex” or difficult tasks. The difficulty of the tasks is defined by the effort and physical activity required to finish the individual task (number of mouse clicks and wheel turns). A sample set of five tasks is listed in Table 1.

TABLE I. SAMPLE SET OF TASKS

Atomic tasks	
Task 1	Change the fan speed to “3”.
Task 2	Play the song “Yesterday” by The Beatles.
Task 3	Check your fuel level.
Complex tasks	
Task 4	You want to travel to New York City. Please check the traffic report for New York City and tell the name of the street mentioned in the report to the experimenter.
Task 5	Please set your navigation system to take you to New York City. When asked, verbally enter (dictate) the name of the street suggested by the traffic report into the navigation system.

**F. Logged Data & Questionnaires**

1) Driving errors: The driving errors and anomalies are recorded by the driving simulator software and saved into a database. The records in the database contain the information about the occurrence of an error, its description, and severity. The later is described using penalty points. The ML machine is responsible for organizing and archiving the driving data for each subject per experimental condition.

2) Task Completion Times: Task completion times are measured manually by the operator of the ML machine using a logging application. The measurement starts when the instruction “Please start now!” is given to the test subject and it is completed when the task is completed successfully.

3) Video Recordings: The entire user study is recorded with a HDR-XR105 Sony digital video camera. The recordings are used to perform an additional post-evaluation of driving performance and general safety. This allows for a recap and further analysis of situations in which severe driving errors were recorded by the system.

4) NASA TLX: Hart and Staveland’s NASA Task Load Index (TLX) method assesses work load on five seven-point Likert scales [15]. It is a subjective, multidimensional assessment tool that rates perceived workload on six different subscales: Mental demand, physical demand,

temporal demand, performance, effort, and frustration. An example of the TLX is given in Figure 5.

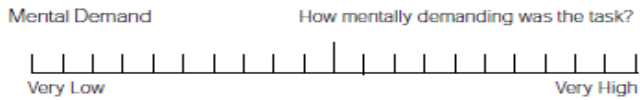


Figure 5. Example of items in the NASA Task Load Index.

5) User Experience Questionnaire (UEQ): The UEQ [16] is a tool for the user-driven assessment of software quality and usability. It consists of 26 bipolar items, each to be rated on a seven-point Likert scale. It has been developed to measure six factors: Attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. An example of three items is presented in the Figure 6.

Annoying	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	enjoyable
Predictable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unpredictable
Efficient	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	inefficient

Figure 6. Example of items in the User Experience Questionnaire.

### G. Experiment Procedure

Before the experiment participants are given a thorough introduction to the driving simulator, the interaction device, and the menu structure. After they familiarize themselves with the simulator and the menu (approx. 20 minutes) they begin the experiment by either first using the visual, the audio-visual, or the auditory interface depending on their random assignment to one of six groups. During the experiment, participants are asked to drive either on a motorway or through a busy city center while they are given first a set of atomic tasks followed by a set of complex tasks. After each condition, participants complete the electronic version of the NASA TLX followed by a complete UEQ to evaluate their experience of the particular interface they just used. After participants complete all three conditions they are asked to fill a short post-study questionnaire on their overall perception of the interfaces, the readability of the projection, the sound design, the realism of the driving simulator, and the task design.

## IV. PRELIMINARY RESULTS

A total of 30 test subjects participated in the experiment. The proposed experimental setup and apparatus have proven to be robust, flexible, and suited for evaluating the interfaces.

The brief analysis of task completion times identified the audio-visual interface as the fastest and the audio interface as the slowest of all three.

On the other hand, the best control of the car was noticed when using the audio interface and the worst when using the visual interface. We also noticed a higher compliance with the traffic rules when the test subjects were performing the tasks compared to just driving. This could partly be explained by a reduction in driving speed when performing the tasks.

Based on the preliminary results, the participants show an

overall preference for the audio-visual combination. However, one fourth prefers only the audio or the visual interface respectively. The thorough analysis of all data collected in the experiment will bring more detailed insights that will enable further adjustments and user testing of the proposed interfaces.

## REFERENCES

- [1] D. L. Hibberd, S. L. Jamson, and O. M. J. Carsten, "Managing in-vehicle distractions: evidence from the psychological refractory period paradigm" Proc. of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 4-11, 2010.
- [2] M. A. Pettitt and G. E. Burnett, "Defining driver distraction," Proc. of World Congress on Intelligent Transport Systems, San Francisco, USA, 2005.
- [3] S. Damiani, E. Deregis, and L. Andreone, "Driver- vehicle interfaces and interaction: where are they going?" European Transport Research Review, vol. (2), pp. 87-96, 2009.
- [4] C. D. Wickens, "Processing resources in attention," in Varieties of Attention (R. Parasuraman and R. Davies, eds.), pp. 63-102, New York, NY, USA: Academy Press, 1984.
- [5] P. Green, "Crashes induced by driver information systems and what can be done to reduce them," in Proceedings of Convergence 2000, Soc. of Automotive Engineers, (Warrendale, PA, USA), pp. 26-36, Society of Automotive Engineers, 2000.
- [6] W. Wierwille and L. Tijerina, "Vision in vehicles vi," in Modelling the Relationship Between Driver In-Vehicle Visual Demands and Accident Occurrence (A. Gale, I. Brown, C. Haslegrave, and S. Taylor, eds.), pp. 233-244, Elsevier, 1998.
- [7] K. L. Young, M. A. Regan, and M. Hammer, "Driver distraction: A review of the literature," Tech. Rep. 206, Monash University Accident Research Centre, Victoria, Australia, 2003.
- [8] L. Yung-Ching, "Effects of using head-up display in automobile context on attention demand and driving performance" Displays, vol. 24(4-5), pp. 157-165, 2003.
- [9] E. Fisher, R.F. Haines, and T.A. Price, "Cognitive issues in head-up displays," NASA Technical Paper 1711, NASA Ames Research Center, Moffett Field, CA, 1980.
- [10] C. D. Wickens, R. Martin-Emerson, and I. Larish, "Attentional tunneling and the head-up display," in: R.S. Jensen (Ed.), Proceedings of the Seventh International Symposium on Aviation Psychology, Ohio State University, Columbus, pp. 865 - 870, 1993.
- [11] J. Sodnik, C. Dicke, S. Tomazic, and M. Billingham, "A user study of auditory versus visual interfaces for use while driving," Int. J. Human-Computer Studies, vol. 66(5), pp. 318-332, 2008.
- [12] M. Jeon, B. K. Davison, M. A. Nees, J. Wilson, and B. N. Walker, "Enhanced auditory menu cues improve dual task performance and are preferred with in-vehicle technologies," Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 2009.
- [13] J. D. Lee, B. Caven, S. Haake, and T. L. Brown, "Speech-based inter- action 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(4), pp. 631-640, 2001.
- [14] City Car Driving, Car Simulator, <http://citycardriving.com/>, accessed 14. November 2011
- [15] S. G. Hart, "NASA-Task Load Index (NASA-TLX); 20 years later," in Human Factors and Ergonomics Society 50th Annual Meeting San Francisco, CA, pp. 904-908, 2006.
- [16] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in: Holzinger, A. (Ed.): USAB 2008. LNCS 5298, pp. 63-76, 2008.

## Evaluating the Usability and the Communicability of Grid Computing Applications

Cristian Rusu, Silvana Roncagliolo,  
Arturo Figueroa  
Pontificia Universidad Catolica de Valparaiso  
Valparaiso, Chile  
e-mail: cristian.rusu@ucv.cl, silvana@ucv.cl,  
arturo.figueroa.r@mail.pucv.cl

Virginica Rusu  
Universidad de Playa Ancha  
Valparaiso, Chile  
e-mail: virginica.rusu@upla.cl

Dorian Gorgan  
Technical University of Cluj-Napoca  
Cluj-Napoca, Romania  
e-mail: dorian.gorgan@cs.utcluj.ro

**Abstract**—Usability is a main attribute of any interactive software system. Its relevance for Grid Computing applications is expected to increase, as the technical knowledge of grid users will gradually decrease. Usability evaluation for Grid Computing applications brings new challenges. A set of specific usability heuristics was defined and validated. The paper presents a Grid Computing communicability study and evaluates the communicability's impact on applications' usability.

**Keywords**—usability; communicability; grid computing applications; semiotic engineering; usability heuristics.

### I. INTRODUCTION

Grid computing is a relatively new, distributed computing technology, which relies on the coordinated use of different types of computing resources of an unspecified number of devices. The ISO/IEC 9241 standard defines the usability as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use [1]. It is expected that the technical knowledge of grid users will gradually decrease, therefore the usability of Grid Computing applications will soon become a main issue.

Usability evaluation for applications based on emerging information technology brings new challenges. Is it the classical concept of usability still valid? Which are the dimensions of the (new) usability? How can it be measured? How should we develop for (better) usability? There is a need for new evaluation methods or at least for the use of traditional evaluations in novel ways [2].

The communicability is defined as the distinctive quality of interactive computer based systems that communicate efficiently and effectively to users their underlying design intent and interactive principles [3]. Communicability has (potentially) a major impact on system's usability.

A set of 12 usability heuristics for Grid Computing applications and an associated usability checklist were defined and validated in several case studies [4]. Later on, a

semiotic inspection was performed, in one case study. The paper explores the communicability's impact on applications' usability. Section 2 summarizes the usability heuristics proposal for Grid Computing applications. Section 3 shows the results of the semiotic inspection and highlights the relationship between application's usability and communicability. Conclusions are presented in Section 4.

### II. USABILITY IN GRID COMPUTING APPLICATIONS

Grid Computing users, their knowledge and specific tasks may be categorized as follows [5]: (1) service end-user, (2) service end-user execute, (3) power user agnostic of grid resource nodes, (4) power user requiring specific grid resource nodes, (5) power user developing a service, (6) service provider, (7) infrastructure system administrator. It is expected that the technical knowledge of grid users will gradually decrease. The number of users belonging to the first and the second of the above mentioned categories are growing fast. That is why we think the usability of Grid Computing applications will soon become a main issue.

Heuristic evaluation is a widely used inspection method [6]. A group of evaluators inspect the interface design based on usability principles (heuristics), usually Nielsen's ten heuristics [7]. Heuristic evaluation is easy to perform, cheap and able to find many usability problems (both major and minor problems). However, it may miss domain specific problems. That is why the use of appropriate heuristics is highly relevant.

In order to develop specific usability heuristics for Grid Computing applications, a 6 steps iterative methodology was followed [8]. A set of 12 usability heuristics and an associated usability checklist of 42 items were defined [4]. The 12 heuristics were grouped in three categories: (1) *Design and Aesthetics*, (2) *Navigation* and (3) *Errors and Help*.

The set of 12 new Grid Computing usability heuristics were checked against Nielsen's 10 heuristics, using *GreenView* and *GreenLand* as case studies [9], [10], [11].



Grid Computing heuristics worked better than Nielsen's heuristics, in both cases [4].

### III. COMMUNICABILITY IN GRID COMPUTING APPLICATIONS

The *Special Interest Group on Computer-Human Interaction (SIGCHI)* of the *Association for Computing Machinery (ACM)* defines *Human-Computer Interaction (HCI)* as the discipline concerned with the design, evaluation and implementation of interactive computing systems for human use, and with the study of major phenomena surrounding them [12].

The *Semiotic Engineering* views HCI not just as a communication between users and software systems, but as a computer-mediated communication between designers and users, at interaction time. The system is therefore the designer's deputy, the artifact that transmits designer's intentions [13].

#### A. Semiotic Engineering and Communicability

The semiotic engineering considers HCI as an interactive and progressive communication process about how to communicate with the system, when, why, and to what effects. The software system speaks for its designers in various types of conversations, specified at design time. The process is one of communication about communication, or metacommunication. Communicability is the attribute that defines the quality of the metacommunication.

The content of the designer message can be paraphrased by a generic metacommunication template: "*Here is my understanding of who you are, what I've learned you want or need to do, in which preferred ways, and why. This is the system that I have therefore designed for you, and this is the way you can or should use it in order to fulfill a range of purposes that fall within this vision*" [14].

There are three distinctive classes of signs in the designer's deputy's interactive discourse:

- *Static signs*, whose meaning is interpreted independently of temporal and causal relations;
- *Dynamic signs*, which are bound to temporal and causal aspects of the interface, emerging with the interaction;
- *Metalinguistic signs*, which explicitly communicate to users the meanings encoded in the system and how they can be used.

Static signs stimulate the user to engage in interaction with the system; they help the user anticipate what the interaction will be like and what consequences it should bring about. Dynamic signs confirm or disconfirm the user's anticipation. The meaning of static and dynamic signs is explicitly informed by metalinguistic signs.

#### B. Communicability Evaluation

The semiotic engineering offers two methods to evaluate the quality of metacommunication in HCI:

- The *Semiotic Inspection Method (SIM)*, and
- The *Communicability Evaluation Method (CEM)*.

SIM explores the emission of metacommunication, seeking to reconstruct its content, expressions, and targeted receivers. CEM explores the reception of

metacommunication, seeking to identify, by means of user observation, empirical evidence of the effects of the designers' messages as they are encountered at interaction time [14]. SIM is an inspection method, involving only specialized evaluators. CEM is a test method, involving real and/or representative users.

SIM aims to reconstruct designer's message using the metacommunication template as a guide. It includes five core steps:

1. The analysis of metalinguistic signs,
2. The analysis of static signs,
3. The analysis of dynamic signs,
4. A comparison of the designer's metacommunication message generated in the previous steps,
5. A final evaluation of the inspected system's communicability.

In steps 1, 2, and 3, the evaluator does a segmented analysis of the system, deconstructing the metacommunication message. It allows to inspection of how the designer communicates with each type of sign (each main communication channel). In steps 4 and 5, the evaluator reconstructs the metacommunication message (filling out the metacommunication template) by comparing, integrating, and interpreting the data collected in previous steps of the method.

#### C. The Semiotic Inspection of a Grid Computing Application

A semiotic inspection of *GreenView* was done. First, the metalinguistic signs were analyzed. They are limited to pop-up messages explaining functionality (Fig. 1), and some general explanation (Fig. 2).

The metacommunication template based on metalinguistic signs may be synthesized as follows: "*You are an expert user that knows what to do. You need to process large amount of specific data, efficiently, in order to get specific information. I have designed a specialized system, that you now how to use, therefore you don't need explanations*".

Static signs are abundant. They include specific layout, controls grouping, icons, and explicit menu options. Fig. 3 shows the "*Fine-to-Coarse*" control panel, highlighting controls grouping (green color), icons and menu options (red color).

The metacommunication template based on static signs may be synthesized as follows: "*You are a user that knows what to do. You need to get easy access to specific functionality, to do actions efficiently, in order to perform specific processes. I have designed a specialized system, which you should be able to use efficiently*".

Dynamic signs include explicit information on system's status (Fig. 4), alternative options of (intuitive) direct manipulation (Fig. 5), and pop-up windows (Fig. 6).

The metacommunication template based on dynamic signs may be synthesized as follows: "*You perfectly know what the system offers. You need to work efficiently, choosing your own way. I have designed a specialized system, which offers alternative ways and feedback that you should understand*".

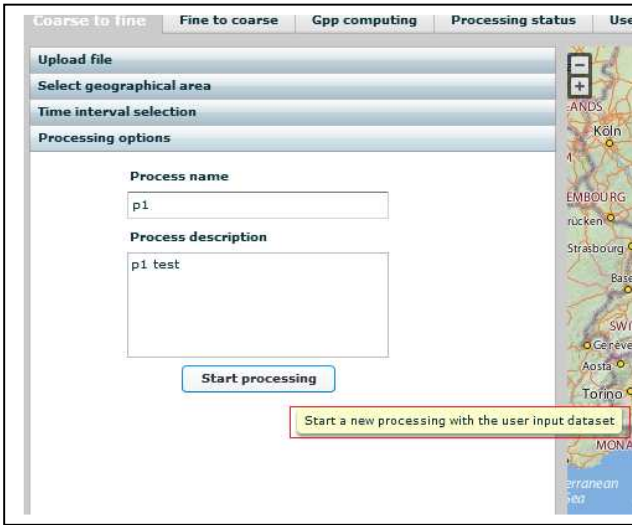


Figure 1. Pop-up messages explaining functionality.

Metalinguistic signs are scarce. Textual static signs are themselves (a kind of) metalinguistic signs.

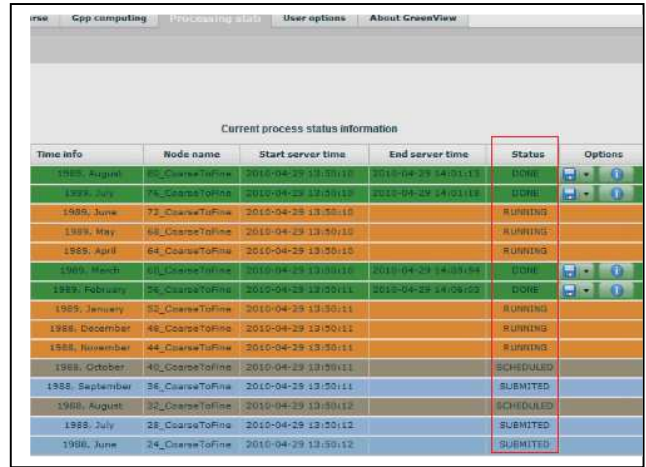


Figure 4. Explicit feedback on processing status.

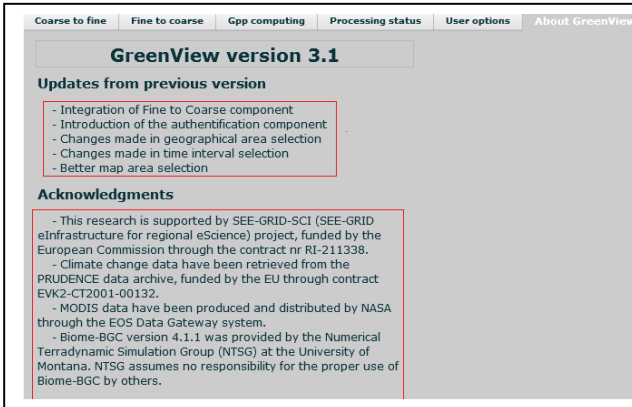


Figure 2. Explanations about GreenView application

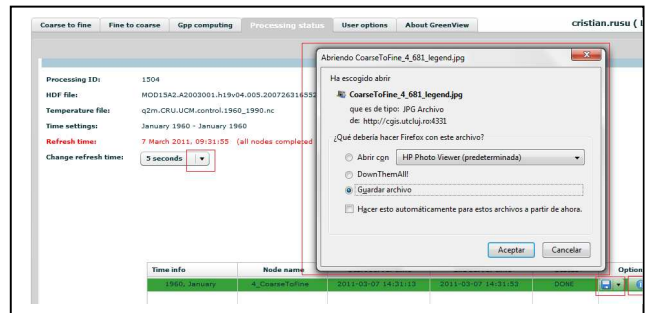


Figure 5. Alternative direct manipulation options.

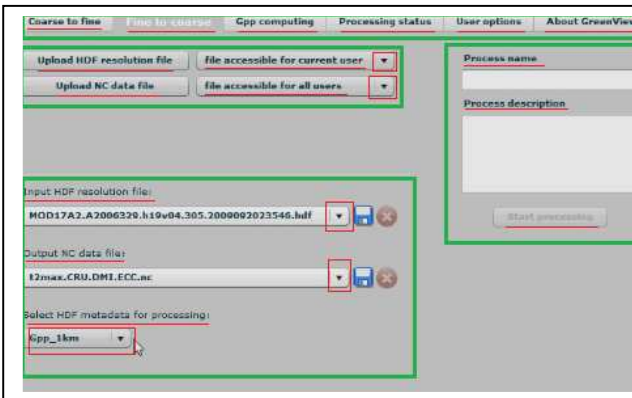


Figure 3. Static signs on "Fine-to-Coarse" control panel.

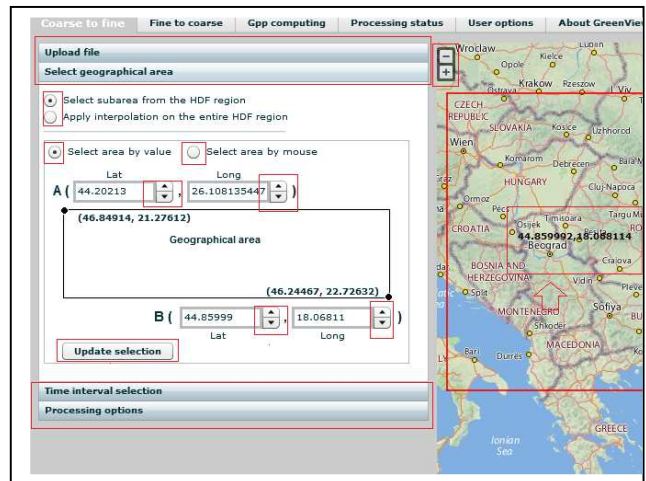


Figure 6. Pop-up windows.

The three metacommunication messages generated by metalinguistic, static and dynamic signs are homogeneous. Communication strategies based on static and dynamic signs are quite similar and coherent. The metacommunication message generated by metalinguistic signs is rather cryptic.

The overall evaluation of GreenView's communicability shows that the system is oriented to expert users. The system is still under development. New functionality is added, in new system's versions. Efficiency and flexibility are explicit system's goals.

D. Usability and Communicability in GreenView

Table 1 shows the number of usability problems identified in *GreenView*. A significant amount of problems were associated to *Errors and Help* heuristics: H10 (*Error prevention*), H11 (*Recovering from errors*), and H12 (*Help and documentation*). As the semiotic inspections proved, there is a lack of metalinguistic signs in *GreenView*. Therefore such usability problems were somehow expected.

The metacommunication global message of *GreenView* highlights its focus on expert users. That could explain a relatively large number of usability problems associated to heuristic H1 (*Clarity*). As the metacommunication global message shows, *GreenView* also focus on efficiency and flexibility. These goals seem to be accomplished, as a low number of usability problems were associated to heuristics H5 (*Consistency*), H6 (*Shortcuts*), H8 (*Explorability*), H9 (*Control over actions*), and no usability problems were associated to heuristic H7 (*Low memory load*).

An evident relationship between application’s usability and communicability may be observed. The global metacommunication message highlights application’s goals, users’ profile, and communication strategies, anticipating associated usability problems.

TABLE I. USABILITY PROBLEMS IDENTIFIED IN *GREENVIEW*

Usability Heuristic	Associated Usability Problems
H1: Clarity	3
H2: Metaphors	1
H3: Simplicity	0
H4: Feedback	2
H5: Consistency	1
H6: Shortcuts	1
H7: Low memory load	0
H8: Explorability	1
H9: Control over actions	1
H10: Error prevention	3
H11: Recovering from errors	2
H12: Help and documentation	2
<b>Total:</b>	<b>17</b>

IV. CONCLUSIONS

The current use of Grid Computing is at the hand of experts and researchers, but it is expected that in the future the technical knowledge of grid users will decrease. Research usually focuses on Grid Computing based application development from a technical point of view. There is a need for new usability evaluation methods or at least evaluations should be particularized for Grid Computing environments. A set of 12 Grid Computing usability heuristics and an associated usability checklist were specified, validated and refined by an iterative process.

The semiotic engineering brings a new perspective on human–computer interaction. Semiotic inspections identify the designer–to–user message reconstructing the metacommunication template. Communicability evaluation may anticipate associated usability problems. Therefore it may be a powerful tool, for researchers, usability professionals and Grid Computing application developers.

ACKNOWLEDGMENT

The authors would like to thank to all the participants involved into the experiments that the present study required. The work was highly supported by the *School of Informatics Engineering of the Pontifical Catholic University of Valparaiso* – Chile and by the *Computer Science Department of the Technical University of Cluj-Napoca* - Romania.

REFERENCES

- [1] International Organisation for Standardisation, ISO 9241, “Software Ergonomics Requirements for office work with visual display terminal (VDT)”, Geneva, 1998.
- [2] C. Wiberg, K. Jegers, and H. Desurvire, “How Applicable is Your Evaluation Methods – Really?”, Proc. 2nd International Conferences on Advances in Computer-Human Interactions (ACHI 2009), IEEE Press, ISBN: 978-1-4244-3351-3, 2009, pp. 324-328.
- [3] C. de Souza C., “The Semiotic Engineering of Human-Computer Interaction”, MIT Press, 2005.
- [4] S. Roncagliolo, V. Rusu, C. Rusu, G. Tapia, D. Hayvar, and D. Gorgan, “Grid Computing Usability Heuristics in Practice”, Proc. 8th International Conference on Information Technology: New Generations (ITNG 2011), IEEE Press, 2011, pp. 145-150.
- [5] M. Norman, “Types of grid users and the Customer-Service Provider relationship: a future picture of grid use”, University of Oxford, 2006.
- [6] J. Nielsen, “Usability Engineering”, Morgan Kaufmann Publishers, 1993.
- [7] J. Nielsen, “Ten Usability Heuristics”, www.useit.com/papers/heuristic/heuristic\_list.html (retrieved: 11/11/2011).
- [8] C. Rusu, S. Roncagliolo, V. Rusu, and C. Collazos, “A Methodology to Establish Usability Heuristics”, Proc. 4th International Conferences on Advances in Computer-Human Interactions (ACHI 2011), IARIA, ISBN: 978-1-61208-003-1, 2011, pp. 59-62.
- [9] R. Meszaros, I. Lagzi, Z. Barcza, and G. Gelybo, “GreenView Application Specifications”, Eotvos Lorand University, Hungary, 2008.
- [10] D. Gorgan, T. Stefanut, V. Bacu, and D. Mihon, “Grid based Environment Application Development Methodology”, LNCS 5910, Springer, 2010, pp. 499-506.
- [11] C. Rusu, S. Roncagliolo, T. Gonzalo, H. Danae, V. Rusu, and D. Gorgan, “Evaluating the Usability of Intercultural Collaboration Platforms: Grid Computing Applications”, Proc. 3rd ACM International Conference on Intercultural Collaboration (ICIC 2010), ACM Digital Library, ACM 978-1-4503-0108-4/10/08, 2010, pp. 179-182.
- [12] ACM’s Special Interest Group on Computer-Human Interaction, www.sigchi.org (retrieved: 11/11/2011).
- [13] H. Sharp, Y. Rogers, and J. Preece, “Interaction Design: Beyond Human-Computer Interaction”, Wiley, 2007.
- [14] C. de Souza and C. Leitão, “Semiotic Engineering Methods for Scientific Research in HCI”, Morgan & Claypool, 2009.

## Practical Usability in XP Software Development Processes

Zahid Hussain<sup>\*‡</sup>, Martin Lechner<sup>†</sup>, Harald Milchrahm<sup>†</sup>, Sara Shahzad<sup>†</sup>, Wolfgang Slany<sup>†</sup>,  
Martin Umgeher<sup>†</sup>, Thomas Vlk<sup>†</sup>, Christina Köffel<sup>‡</sup>, Manfred Tscheligi<sup>‡</sup> and Peter Wolkerstorfer<sup>‡</sup>

<sup>\*</sup>*Quaid-e-Awam University, Nawabshah, Pakistan*

<sup>†</sup>*Institute for Software Technology, Graz University of Technology, Austria*

*Email: {zhussain, hmilchra, sshahzad, wolfgang.slany, mumgeher, tvlk}@ist.tugraz.at*

<sup>‡</sup>*CURE – Center for Usability Research & Engineering, Vienna, Austria*

*Email: {koeffel, mtint, wolkerstorfer}@cure.at*

**Abstract**—This paper describes the experiences made and lessons learned in an Extreme Programming (XP) software development project. We investigate the potential of XP to produce user experience-optimized products by including HCI experts in the team. We relate the software development method to user-centered design instruments and propose solutions to different user experience integration problems. Additionally, the practicability of different HCI instruments regarding solving those problems is examined. The analyzed instruments and methods are: user studies, personas, usability tests, user experience expert evaluations, and extended unit tests. The conclusion provides tips and tricks for practitioners.

**Keywords**-Agile Methods; Extreme Programming; Usability; User Experience; User-Centered Design.

### I. INTRODUCTION

As computers and mobile phones have become an essential part of everyone's life, the target audience of software applications has shifted from technical experts to general consumers. Accordingly, the typical usage context of information and communication technologies (ICT) has moved from the office to the home. ICT is used for entertainment in a way that was not present in the days of office-only usage. Therefore, the user experience has become an increasingly important aspect of the adoption and the success of a software product. Today's users expect powerful but nevertheless easy to use applications. Hence, user-centered design (UCD) processes are needed.

At the same time the pressure on the software market has increased. Shorter time-to-market as well as the constantly changing needs of the market and customers impose new demands to software development processes. Agile lightweight methods like Extreme Programming (XP) are designed to cope with these new conditions. For the development of a successful software product it is therefore inevitable to integrate user experience into an agile software development process.

In this paper, we provide background information and advice based on the experiences made in a research project with a practical application. The goal of the project was to increase the quality of software development processes.

Although we mainly discuss agile development processes (specifically XP), we also provide relevant information on how to include user experience techniques into other existing development processes.

### II. BACKGROUND

#### A. Software Development Processes

Software development processes are an attempt to structure and standardize development to make the outcome of a project plannable and predictable. There are many process models available and being used, ranging from heavyweight to agile. In accordance with the project circumstances the actual development process has to be tailored to the specific needs.

The choice of the right process depends on various factors: Risk level, requirements stability, time-to-market, etc. The clearer the requirements and the more stringent the organizational structure, the more structured and heavyweight the process has to be. The more unstable the requirements are, the more iterative and agile the underlying process has to be to cope with the changes. The shorter the time-to-market, the more lightweight and iterative the process should be to avoid administrative overhead and provide deployable software after each iteration [1].

#### B. Agile Software Development

The next logical step was the invention of agile software development methods. In software engineering, agile software development or agile methods refer to low-overhead methodologies that accept that software is difficult to control. They minimize risk by ensuring that software engineers focus on smaller units of work, business priorities, and high quality. One way in which agile software development is generally distinguished from "heavier", more process-centric methodologies, for example the waterfall model, is by its emphasis on values and principles rather than on processes. Typical development cycles are one week to one month. At the end of each cycle the project priorities are re-evaluated. This is a feature that is shared with iterative

development methodologies and most modern theories of project management.

1) *Extreme Programming*: One well known method from the group of agile software development methods is XP. The XP methodology was formulated by Kent Beck, Ward Cunningham and Ron Jeffries. In March 1996 Kent Beck started a project at DaimlerChrysler using new concepts in software development [3]. The result was XP. The starting point was to find out what made software easy to create and what made it difficult. Kent Beck came to the conclusion that there are four factors to improve a software project: communication, simplicity, feedback, and courage are the values sought out by XP programmers [3]. In the second edition of his book, Beck added 'respect' as the fifth value [4].

### III. USER EXPERIENCE AND SOFTWARE DEVELOPMENT

The success of a software development project is not only associated with tools and technologies but also depends on how much the software development process helps to be user-centered and end-user-oriented [5].

#### A. User experience in XP/Agile Projects

Some experts doubt that the XP process leads to true user-centered design [6]. The issues arising from this problem statement suggest that XP and UCD would not fit. But this perception is simplistic and misguided as shown by successful applications in practice [7]. We can see succeeding practitioners combining user experience/UCD and XP/agile methods by varying approaches [5][7][8][9][10][11][12][13][14][15][16][17][18][19][20][21][22][23][24].

1) *Why do some XP projects fail when including HCI work?*: The following issues can prevent the integration of HCI instruments into XP processes [17]:

- Ad-hoc Input: Because of the short release cycles software engineers would need ad-hoc user experience input during development. In practice, user experience input is not given ad-hoc, but after longer periods (one to two weeks average). Such time-spans are not acceptable for most XP practitioners.
- Cultural problems: Software engineers on the one hand and HCI experts on the other hand come from different domains with different attitudes, approaches, backgrounds, and even different ways to express themselves while communicating. The XP process requires tight cooperation in teams, which reveals differences between engineers and HCI experts very quickly: engineers have a technical approach to software development whereas HCI experts mainly have psychology backgrounds, hence taking a cognitive view on software development. As these differences can lead to problems, methods to prevent this have to be integrated into the collaboration process.

- Technical Focus: By its genesis unit tests in XP environments are designed for technical testing. Hence, the focus is on technical functionality – ignoring user experience issues. This means that the technical view of testing has to be expanded by HCI approaches and means.
- On-site Customer Representative: From an HCI point of view the inclusion of customers is a step into the right direction. But the Manifesto for Agile Software Development does not clearly demand end-users as customers [49]. We expect deficits in user experience if it is not clearly stated that end-users have to be part of the process. Developers need to have a clear picture of the humans they develop for.
- Awareness: In order to successfully include user experience and user-centered design in an XP process the developers need to have a basic understanding of user experience issues.

### IV. CASE STUDIES

#### A. Setting

Mobile computing is leading a revolution. Studies show that multimedia – Audio and Video (AV) – consumption is on the edge to become one of the next killer applications for mobile devices [25]. Still, there are not many full-featured applications on the market which utilize the available bandwidth and are accepted by consumers.

We have been developing an application in this field that enables a user to perform content-based search for AV content in large digital archives and play it on a mobile phone. The major problem for an average user in this context is the combination of the overwhelming amount of multimedia content available and unsatisfactory user-interfaces for accessing it. User experience is the key success factor for such applications [5][26].

#### B. Approach

The novelty of our approach lies in the fact that we do not only use one or two methods related to user experience to integrate them into the XP process but selected five instruments to enhance the existing XP process. This multi-instrument approach was developed to solve the problems as introduced in section III-A. Applied correctly in different phases of the project these instruments are designed to reach the goal of maximized software quality in terms of technical quality and also in terms of user experience [5][17][18]. The five instruments we relied on are:

- Extreme Personas
- User Studies
- Usability Tests
- User Experience Expert Evaluations
- Extended Unit Tests

The integration of these HCI instruments into the XP process is shown in Figure 1. The following sections will provide a short overview of the single methods.

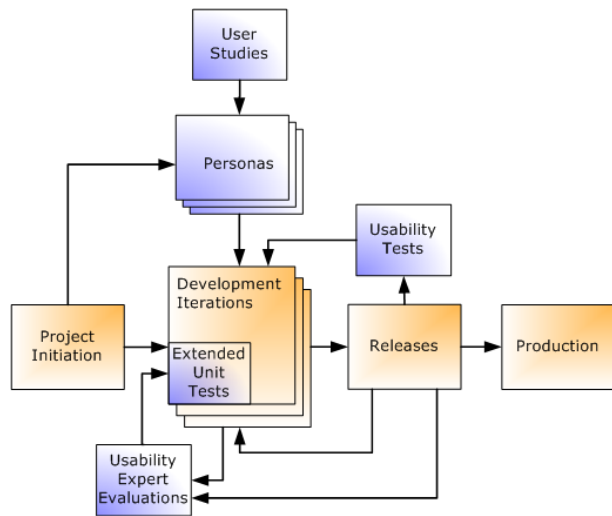


Figure 1. The modified agile development process with user experience instruments included (Extreme Personas, User Studies, Usability Tests, User Experience Expert Evaluations, and Extended Unit Tests). End-users are integrated in two different ways into this process: on the one hand user studies inform the development and extension of the personas (which indirectly provides input to the developers); on the other hand usability tests (as a part of the user experience evaluations) directly inform the development [5][17].

### C. Personas

“Personas are not real people, but they represent them throughout the design process. They are hypothetical archetypes of actual users. Although they are imaginary, they are defined with significant rigor and precision.” [27, pp. 123-124].

Personas are a design tool based on the ideas of Alan Cooper, who released his book “The Inmates are Running the Asylum” in 1999, which is considered to be the founding work in the field of personas [27]. Since the invention of personas, many scientists but also large companies have gathered interest in this approach.

The personas method was developed as a tool for rising empathy for the end users in development teams and as a means for communicating peer group definitions [17][18]. Personas determine what a product should do and how it should behave. They communicate with stakeholders, developers, and other designers. Furthermore, personas build consensus and commitment to the design and measure the design’s effectiveness. They also contribute to other product-related efforts such as marketing and sales plans [27].

They describe the target user – his wishes, desires and application-specific aspects. Furthermore, they show the nature and scope of the design problem [28]. If no personas are defined in a project, the project members will always

envision themselves to be the end-user, which leads amongst others also to communication problems [28]. Each persona represents a peer group of users. With a detailed description of the personas, every member of the project knows the main users and has a unified view of the target customers [27][28].

The advantages of personas are that they allow to unify the picture of the target user for the project team, which allows for a more fluent communication [28]. Additionally, personas make use of the “emotional mind” [29] of people which leads to a better focus on user-centered thinking within a project. “The user” as an abstract term is eliminated from the project and is replaced by people with names and faces, which also saves time because of, e.g., shortened debates. With the personas method, no existing processes have to be modified or changed; personas are just added to the project to focus more on the end-users.

Moreover, personas allow for informed design and according to Alan Cooper, they enlighten the design process [27]. Furthermore, personas can also be used as an evaluation tool such as walkthroughs [28]. As an archetypical figure, personas can guide decisions about product features, navigation, interactions, and even visual design (among other factors) [30].

In the agile development process, personas can be integrated as so-called “Extreme Personas” [17], an approach to personas that starts with the same activities as in the classical persona method: preliminary user groups are defined and personas are modeled according to them. In addition, the knowledge gathered in user studies is incorporated and the personas will be refactored when new knowledge suggests slight changes. If the found knowledge reveals that current personas do not cover the insights, new personas will be developed.

These actions make the classical personas “extreme” by applying the XP paradigm of small iterative steps and refactoring – which is extending the personas in this case. During the coding phases the developers pin the personas beside the user stories. Their first application is in the planning games (the phase where user-stories are created) where Extreme Personas represent the on-site customer.

### D. User Studies

User studies are the instrument for getting knowledge about end-users. The outcome of the user studies informs the design in two ways: on the one hand knowledge for creating and refactoring the personas is obtained; on the other hand direct input for the user stories can be derived [17][18].

In the agile process, user studies were employed in two different ways [18]:

- Laddering interviews and
- Field trials

The following sections provide a short introduction to both of these areas.

1) *Laddering Interviews*: Laddering interviews are techniques that are mostly applied in the field of marketing and psychology. Nevertheless, recently they have also been applied to investigate user experience [31]. In a structured interview between two persons, the connections between attributes and their consequences are investigated from the interviewee's point of view [32].

The duration of a laddering interview can be – depending on the content – between forty-five minutes and two hours. Structured questions are employed to discover the respondent's beliefs, feelings, and goals. In order to get familiar with the interviewee, a warm-up phase is needed and sessions are usually recorded.

2) *Field Trials*: In field trials the developed products are tested by real users in an uncontrolled setting. The feedback of the users can be collected by applying various techniques. These techniques may include surveys, questionnaires, interviews, contextual inquiries, diary studies, etc.

#### E. Usability Tests

Usability tests are empirical studies that involve real users [18]. They can be considered as the most fundamental user experience evaluation method [33]. Contrary to field trials, usability tests take place in laboratory environments. Usability tests are conducted several times during product development to measure accuracy, user performance, recall-value, and the user's emotional response.

During the tests the users are observed using the product. In some cases the users can be asked to think aloud and verbalize their thoughts to get a better insight to the user's mental model, as well as encountered problems. Other methods such as interviews can be combined with usability tests.

In our project usability tests in the laboratory have included end-users as demanded by the UCD process but not demanded by the XP process where it is not mandatory that end-users are part of the on-site customer representative [17].

#### F. Expert-based User experience Evaluations

Expert-based user experience evaluations are reviews conducted by experts [18], either in the area of user experience, in the application area of the particular system, or both. The following two expert-based user experience inspection methods are the most renowned:

- Heuristic Evaluation
- Cognitive Walkthrough

1) *Heuristic Evaluation*: Heuristic Evaluation has gained in popularity with Molich and Nielsen's introduction of ten heuristics to a wider audience [34][35]. The original heuristics have later been improved and adapted to different areas [33] and are still frequently employed to evaluate different kinds of systems.

This evaluation method is considered an efficient analytical and low-cost user experience enhancement method,

which can be applied repeatedly during a development process. In general, heuristics can be considered as rules of thumb describing the affordance of a user to a system and are formulated more generally than the rather specific guidelines. They are recognized and established user experience principles.

During a heuristic evaluation three to five experts (one expert at a time) inspect a system according to given heuristics. The found issues are categorized according to their severity after the evaluation is finished. Heuristics do not cover all possible occurring user experience problems but because of the ease of application they can be employed very early in the design process – even when usability testing would not be possible.

2) *Cognitive Walkthrough*: A cognitive walkthrough is an analytical user experience inspection method which was introduced by Wharton, Rieman, Lewis and Polson [36]. The main goal of a cognitive walkthrough is to measure the learnability of a system by detecting user experience issues. Traditionally this evaluation method is conducted either by a single expert or a group of experts and novice users who put themselves in the place of a hypothetical user [34]. During the evaluation, typical tasks are accomplished within the four phases of a cognitive walkthrough. Similar to heuristic evaluations, cognitive walkthroughs can be applied very easily and early in the design process.

3) *Application in Agile Environments*: In projects involving XP, expert-based user experience evaluations solve the problem of ad-hoc input. This is done by IM (instant messaging), email, and (video) conferencing. Mock-ups (in early phases) and screens (in later phases) are sent to the HCI experts who then give ad-hoc input by using the mentioned channels [17][18]. For this purpose the methods have to be tweaked in a way to be less time and cost intensive. This is done by involving less experts or users (only 1-2) than recommended. These tests are done much more frequently (every 1-3 iterations) and therefore the results can be accumulated. Important here is to switch users and experts frequently to achieve similar results.

#### G. Extended Unit Tests

Extended unit tests originate from automated usability evaluation (AUE) [18]. The idea of automated usability evaluation is not new. Basic research goes back to the early nineties [37][38]. In the year 2000 the state of AUE was still described as “quite unexplored” [39].

AUE offers usability support through specialized tools. Therefore, developers can be supported by automatic inspection throughout the development phase of a project. An example of such kind of automatic evaluation is log-file analysis [40][41]. Here the generated data helps identify paths and execution time in order to detect problems.

Another example for AUE is the NIST-Web-Suite [50] that allows for an automatic code-based analysis of websites

according to 12 design guidelines. The WAUTR-project [51] (Automatic Usability Testing enviRonment) is a first attempt to support user experience experts with a set of different tools.

The availability of user-generated data already during the development is one problem of AUE. The simulation of the final users [42] or specific aspects (e.g., gaze [52][53]) is one possibility to solve this issue. Nevertheless, also this approach requires user-data.

When no users are available, code-analysis is suggested as the next best solution. The code is used to calculate usability factors and give input on usability based on design guidelines (e.g., the ratio between text and graphics on a website, the number of links, the use of colors, etc.) [39][43].

This approach is currently mainly used in the web area. WebTango is one of the best known tools in this field of application [44]. It calculates usability metrics on basis of HTML code, and the evaluation is based on a statistical model of the website usage.

An approach that goes even further was developed by Abdulkhair [42] who implemented agents that are able to learn from user-behaviors. The underlying statistical model allows the agents to detect user preferences, learn them, and use them to evaluate websites.

The reverse-engineering of the structure of a website was used by Paganelli and Paterno [45] in order to find potential usability problems with their tool “WebRemUSINE”.

Currently, the main target group of such kind of tools are experts in the area of human-computer-interaction. The transition towards developer-based tools for AUE is currently in progress [17]. These tools would then be able to continuously check for user experience issues even during development while the code is being written. Although gaining in popularity, automatic usability evaluation can rarely be found in commercial development environments. The most renowned product of this kind is LIFT [54], which is comparable to WebSAT. Additionally, LIFT integrates the GoLive, FrontPage, and Dreamweaver development environments.

Since most of the current approaches tend to focus on multimodality [46] and mobile devices [47], the existing tools have a big disadvantage for our project: most of them are isolated solutions which are – as mentioned before – solely designed for HCI experts. Hence, they hardly integrate seamlessly into the existing development processes.

In XP, unit testing is mandatory. Our approach extends the technical unit tests by adding usability-specific test cases. Code based tests are enhanced with semantics to achieve this goal – for example: code based tests can check against guidelines like the usage of capital letters on buttons. When adding semantics (the correct label of a button), we can include the test into the set of unit tests already used in XP. Test-driven development in XP means to write tests first. The written tests then define the behavior of the application.

Adding usability related unit tests with semantics allows us to define the user experience of the application. Unit tests – by definition – test small definable units of the software. The problem of patchwork application suggests using a holistic approach to testing. Therefore, unit tests are extended by tests that go beyond single units and test complete interaction flows [17][18].

## V. CONCLUSIONS AND LESSONS LEARNED

We used the previously described process since summer 2007 in our project which ended in summer 2010. The final usability tests, diary studies, focus groups, and the results of log file analysis in the field trial held from December 2008 till May 2009 show that the process is able to really enhance user experience of XP-style developed applications [48].

The tight coupling of different expertise has led to high motivation among project members. Developers gain insight into the subtleties of UCD and HCI experts learn to understand the origin of user experience problems. Especially the diverse technical testing frameworks demand technically aware HCI experts. In practice, this could become a problem when the chosen framework is complex and little time for learning is available [5][17][18].

We have experienced that the inclusion of UCD in the software development processes underpins past experiences: no matter if it is a classical “waterfall” development process or XP – the inclusion of UCD mainly depends on the user experience-awareness and on the mindset present in the project – not on the software development model.

Furthermore, we found that especially ad hoc input can be given sufficiently via mail since most of the time no synchronous communication between the project members is needed. The geographical distance between HCI practitioners and developers can partly be overcome by using phone- or video-conferencing [18].

Also from a customer point of view the communication with a user experience engineer can be done most of the time by e-mail and exchanging mock-ups. Especially for mock-ups it turned out to be important that the user experience engineer actually sees them rather than getting them described. The response times of these methods are usually short enough since for user experience input in the story-writing process it is sufficient to get results within 3 to 4 days. For quick feedback on user experience issues during development or for “urgent” re-planning, user experience engineers should be readily available for a quick advice via cell-phone or chat [18]. However we also had good experiences with on-site visits. This allows the user experience engineer to see in which environment the application is developed and to get a better understanding of the developers and customers.

In our case, the creation of user-stories is supported by HCI knowledge derived from studies, literature, and usability



tests. During our project the story-writing was done mainly by the customer and the technicians prior to planning.

A technician of our team paired with the customer in order to create simple and precise stories for the discussion in the planning game. This pairing forces the customer to explain his expectations in detail. The technician helps to refine the ideas by asking questions which could determine implementation details and gives feedback on the answers. Together they ensure that the story is written in a way which is unambiguous and understood by both sides. This way a lot of time and energy, which would have been spent during team discussions, is saved during the planning game. If the story contains user interface (UI) aspects, the user experience engineer is also included early into the story creation-process. The advantages are timesaving, more motivation (less chances of rework due to preclarified user experience issues), and gaining better understanding of needed usability input early in the development [18].

Continuous monitoring, evaluating, and testing of the UI, and quick intervention can lower the danger of a patchwork-experience. Additionally, we could see that cultural problems between HCI and development seem to depend more on the involved persons than the methods used, and we did not experience the problems reported in the literature.

#### A. Review of the Used Instruments

During this project we have extended traditional XP methods with knowledge derived from different HCI instruments. In the following sections, we provide a review of some the methods employed:

1) *Extreme Personas*: The personas method should enable an end-user focused mindset to be established very quickly and hence should solve the problem of the development focus on the technical part. Additionally, the personas should help to orient the project towards on-site customer AND end-user [18]. During our project we concentrated very much on the customer centered design process as well as the design process itself and nearly neglected the personas.

There have been several issues and discussions on Extreme Personas during our project. First, the initially developed personas were not satisfactorily distributed to either the development team or the customer-on-site by the user experience engineers. Second, the development team did not give much credit to the two personas which were provided [18].

From the point-of-view of the development team and the customer-on-site the main cause for these issues was that especially one persona was so funny that they did not take it serious – nevertheless, they got in touch with them instead of neglecting them fully [18].

*What can be learned out of this?* First, Personas should be introduced like new team members, as they will accompany the whole team during the development process. Therefore, they need “room” and “positive energy”: they should give the

developers and customers a feeling of producing something valuable for someone who they like. Of course, fun can (and should) play a major role within teams – but be careful not to mob a persona!

Second, many technicians think that it does not matter for whom they develop. The design is supposed to be the design-department’s decision, the scope of implemented features is part of the customer’s work, and over-all everything is pre- and post-tested by the user experience engineer anyhow – so why worry? If this would be true, why did all technicians in our development team take part in a months-lasting discussion about the user-group we were developing and producing for? Especially the customers thought that the technical or business-related decision processes are colored by conscious and unconscious inputs. That is why personas should be present in an appropriate form and should have their own stable place around the developers. Stories and features should be developed for the personas and their names should be used on the story cards and during discussions. This will help them stay alive and influence the team.

Therefore, the advice from the customer-side is to make personas available and visible, take part in the process of developing them, introduce them to the team, and have them (consciously!) in mind when planning or undertaking any decision process.

2) *Expert Evaluations*: During our project, it took some time until the customers embraced the possibility of asking the advice of the user experience engineer in advance. This might have been caused by the increasing trust in the user experience engineer over time. After a while the customers have reportedly valued the user experience engineer’s input higher. Another reason was the improved planning method which allowed to prepare stories early enough to have time for user experience input.

The experience with expert evaluation was a very positive one if the results arrived on time. For an XP project the usual way expert evaluations are done is not ideal. Instead of big, long lasting, application wide evaluations, what is needed are smaller and faster evaluations on the story level. If the evaluation result comes after the story or iteration completion, the likelihood of ignoring the input increases dramatically. The reason is that either the input is already outdated due to the quick changes in an agile project, or other stories are higher prioritized by the customer. Consequently, a stripped down, quicker version of the usual expert evaluation process is needed.

*What can be learned out of this?* First: Prepare the user stories at least three days before the actual planning. During this preparation when the story-cards are written, all the mock-ups should be drawn as well (use drawings by hand or simple drawing tools). Then the mock-ups are sent to the user experience engineer for feedback [18].

Feedback from the user experience-side should be quick (maximum 3 days for one week iterations). The advantage:

when introducing the story – it is already usability tested. When technical questions arise during the implementation, for instance that a certain demand from the user experience side would cost too much, it is advised to call the user experience-engineer and have a short (video) conference during planning or whenever this occurs. Far less user experience-fixes are the results of this practice [18].

Thus, it is our advice to involve available user experience engineers into the planning process as early as possible. Use his time and input only when the implementation of the results is immanent or when it is critical for the development. Be careful: Do not shift the evaluation and fixes to “when you have time” because this will never occur and thus no serious user experience input will be realized.

3) *Usability Tests:* We noticed that the usability tests had impact beyond the expected one (which is giving input for the design). We saw that the mindset of developers changed dramatically when seeing real users handling the application during these tests [5][18]. We saw that developers who attended the usability tests got more biased towards user-centered thinking than the others [5]. Some of the developers not having been present wanted to watch the recordings of the tests but did not find the time to do so.

When it comes to the results of the tests there was an agreement that the tests were too early in the project to tell us a lot about the user experience problems of the application. Furthermore, the reporting period was too long. When the report arrived there already had been so many changes in the application that many recommendations were obsolete [18].

Therefore, we would recommend smaller tests after every few iterations of the application (better: for every iteration). To keep costs low and if the system is a very fast moving target, not always the entire system should be tested and the number of test users can also be limited to 1-2. This is compensated by the increased frequency of tests resulting in a similar coverage than a big test. Bigger tests should only be made when no major changes are expected and the system is quite stable [18].

4) *User Studies:* We used user studies in the form of laddering interviews and field studies. In autumn 2007 laddering interviews were conducted. From December 2008 till May 2009, a large field trial study was conducted with 150 real end-users spread throughout Austria who used our application on mobile phones.

The users were able to use their devices freely to access multimedia content and did not have any restrictions. They were only asked to fill in questionnaires sent to their mobile phones and reply to certain SMS. The trial study also included diary studies, contextual interviews, laboratory usability tests, and focus groups. [18]. We also logged the actions of the users. The preliminary log files results were presented in [48].

5) *Lightweight Prototypes:* Besides the HCI instruments mentioned above, we also made use of two different types of mock-ups: Low fidelity paper mock-ups, and high fidelity mock-ups if a more detailed clarification was necessary. We got them evaluated by the customer and afterwards sent them to the user experience engineer for additional feedback [18].

## B. Tips & Tricks for Practitioners

Since we have gathered some experience in the combination of XP and HCI centered methods we are providing some tips and tricks in this area.

From the customer’s point of view it is recommended to learn to cope with possible daily occurring misunderstandings between business and technical team members. During the planning process clarification is particularly important.

Furthermore, the user experience expertise should be integrated very early in the planning phase and also during the story-writing process. Invite your user experience engineer from time to time or visit him during user testing. It is quite a lesson watching real users acting with your prototype the first time and giving feedback.

1) *Tips & Tricks for Waterfall-model Projects:* In waterfall-model projects we propose to follow these basic guidelines for including UCD in the development process:

- End-user inclusion: It is important to include end-users as early as possible. It is best to start cooperating already during the requirement phase.
- Pre-requirements phase: as soon as targeted user-groups are known an HCI engineer should do research on existing knowledge of the interaction behavior.
- Prototyping: Use paper-prototyping to gather early feedback from end users. Tests of the paper mock-ups of GUI designs prior to the implementation are particularly important in waterfall-model projects.
- Team building: Include not only end-users but also HCI experts from the beginning. Requirements should be backed by HCI input to ensure that architectural design of a system does not include user experience threats from the ground-up. Some architectures forbid to hide unnecessary complexity from the user which is why HCI input is also important for technology-centered architectural decisions.

2) *Tips & Tricks for Extreme Programming Projects:* Boiled down to a short list we summarize the actual state of experience:

- Usability Test Videos: A highlight video of the test should be created. Highlight videos save time compared to the full video documentation of usability tests and support a user-centered mind-set. The highlight video should be shown to the whole development team.
- Training: HCI engineers should be trained in XP-story writing to be able to deliver their user experience findings in form of user-stories. This saves an additional

step and makes them immediately usable during the planning game [18].

- Story writing: a developer should support the HCI engineer when writing user-stories [18].
- Proper customer and HCI engineer coordination is necessary for the inclusion of the user experience process in the development [18].
- An experienced on-site XP customer can fill-in the technical gap between HCI engineers and developers.
- A short user experience workshop should be held at the start of the project or before the release in which the usability testing phase begins. It serves as a good platform for all XP team members (managers, customers, and developers) to understand the process and importance of usability testing. It also serves as a good kiff-off time for HCI engineers and developers as they have a lot of work to do together in future.
- A pre-plan should be developed by the HCI engineer that states at which detail level of UI design development to use low-fidelity and high-fidelity mock-ups and when to perform which usability testing process. This will help the customer in planning for the iterations. A short meeting of the HCI engineer and the customer before an iteration planning will further help the customer in UI design stories.
- It is the customer's responsibility to make the up-to-date usability tests reports visible to the developers (either on a dedicated usability-tests board or somewhere near the story board).
- As the customer is the one creating and prioritizing the stories it is his duty to also think about and include user experience aspects in his user stories.

### C. Summary & Future Research Directions

We have been able to show that different HCI methods and techniques impact the user experience of XP projects as intended. As we saw, all need tweaking and fine-tuning to perfectly fit the XP process.

For the future we see different directions of research. One is automated usability evaluation (AUE) and its integration into software development processes. Research on testing-frameworks for AUE will be an important next step where HCI experts and developers should collaborate to ensure the final frameworks fulfill requirements of both disciplines. The need for better AUE is obvious: the quickly growing number of custom software products is not matched by a similar growing number of HCI experts. Hence, tools will be needed to support HCI engineers in handling these. As purely code-based AUE approaches are limited to a certain level, more advanced AUE methods have to be developed.

A second research direction for us concerns the need for more in-depth insight into the persona-method. We have to gain more knowledge about the interrelationship between the modeled personas and the cognitive effects on different

developers. One issue is the perception of persona pictures: theory suggests personas should be "likable" – but what do developers like and dislike? Hence, research in the perception of personas should be broadened. Open questions for us are for example: which features of modeled personas support which outcome in the development? What about the influence of different subjective perceptions of a persona? Collaborating with different disciplines – e.g., the game industry, as they know a lot about "character modeling" or media scientists who model characters for TV series – will be necessary to cover these questions.

On the business side we see the need for a more elaborated process on how to include different stakeholders and their input. We assume that the more stakeholders get involved the higher the need for structured inclusion strategies for all stakeholders will become. Research on these inclusion strategies will be necessary to ensure that the input of each stakeholder is treated the right way.

### ACKNOWLEDGMENT

The research herein was partially conducted within the competence network Softnet Austria ([www.soft-net.at](http://www.soft-net.at)) and funded by the Austrian Federal Ministry of Economics (bm:wa), the province of Styria, the Steirische Wirtschaftsförderungsgesellschaft mbH. (SFG), and the city of Vienna in terms of the center for innovation and technology (ZIT).

### REFERENCES

- [1] M. Lechner, "Curid – a software development method for data-driven framework architectures," Master's thesis, Technical University of Graz, Austria., 2005.
- [2] C. Larman and V. R. Basili, "Iterative and incremental development: A brief history," *Computer*, vol. 36, no. 6, pp. 47–56, 2003.
- [3] K. Beck, *Extreme Programming Explained: Embrace Change (1st Edition)*. Addison-Wesley Professional, 1999.
- [4] K. Beck and C. Andres, *Extreme Programming Explained: Embrace Change (2nd Edition)*. Boston: Addison-Wesley, 2004.
- [5] Z. Hussain, M. Lechner, H. Milchrahm, S. Shahzad, W. Slany, M. Umgeher, and P. Wolkerstorfer, "Agile User-Centered Design Applied to a Mobile Multimedia Streaming Application," in *USAB 2008*, ser. LNCS, vol. 5298/2008. Springer Berlin / Heidelberg, November 2008, pp. 313–330.
- [6] W. Hudson, "A tale of two tutorials: a cognitive approach to interactive system design and interaction design meets agility," *interactions*, vol. 12, no. 1, pp. 49–51, 2005.
- [7] P. McInerney and F. Maurer, "UCD in agile projects: dream team or odd couple?" *Interactions*, vol. 12, no. 6, pp. 19–23, 2005.

- [8] J. Patton, "Hitting the target: adding interaction design to agile software development," in *OOPSLA 2002 Practitioners Reports*. Seattle, Washington: ACM, 2002.
- [9] S. Chamberlain, H. Sharp, and N. Maiden, "Towards a framework for integrating agile development and user-centred design," in *7th International Conference on Extreme Programming and Agile Processes in Software Engineering, XP 2006*, ser. LNCS, vol. 4044. Heidelberg, Germany: Springer Verlag, 2006, pp. 143–153.
- [10] A. Holzinger, M. Errath, G. Searle, B. Thurnher, and W. Slany, "From extreme programming and usability engineering to extreme usability in software engineering education (XP+UE→XU)," in *COMPSAC '05: Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSAC'05) Volume 2*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 169–172.
- [11] A. Holzinger and W. Slany, "(XP+UE→XU) praktische erfahrungen mit extreme usability," *Informatik Spektrum*, vol. 29, no. 2, pp. 91–97, 2006.
- [12] D. Sy, "Adapting usability investigations for agile user-centered design," *Journal of Usability Studies*, vol. 2, no. 3, p. 112132, 2007.
- [13] J. Ferreira, J. Noble, and R. Biddle, "Agile development iterations and UI design," in *Agile 2007*. IEEE Computer Society, 2007, pp. 50–58.
- [14] J. Ferreira, J. Noble, and R. Biddle, "Up-front interaction design in agile development," in *8th International Conference on Agile Processes in Software Engineering and eXtreme Programming, XP 2007, Jun 18-22 2007*, vol. 4536 LNCS. Heidelberg, D-69121, Germany: Springer Verlag, 2007, pp. 9–16.
- [15] D. Fox, J. Sillito, and F. Maurer, "Agile methods and User-Centered design: How these two methodologies are being successfully integrated in industry," in *Agile, 2008. AGILE '08. Conference, 2008*, pp. 63–72.
- [16] S. W. Ambler, *Tailoring Usability into Agile Software Development Projects*. Springer London, 2008.
- [17] P. Wolkerstorfer, M. Tscheligi, R. Sefelin, H. Milchrahm, Z. Hussain, M. Lechner, and S. Shahzad, "Probing an agile usability process," in *CHI '08: human factors in computing systems*. New York, USA: ACM, 2008, pp. 2151–2158.
- [18] Z. Hussain, H. Milchrahm, S. Shahzad, W. Slany, M. Tscheligi, and P. Wolkerstorfer, "Integration of extreme programming and user-centered design: Lessons learned," in *XP 2009*, ser. LNBIP, P. Abrahamsson, M. Marchesi, and F. Maurer, Eds., vol. 31. Springer, 2009, pp. 174–179.
- [19] Z. Hussain, W. Slany, and A. Holzinger, "Current State of Agile User-Centered Design: A Survey," in *USAB 2009*, LNCS, vol. 5889. Springer, 2009, pp. 416–427.
- [20] L. Miller and D. Sy, "Agile user experience SIG," in *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems CHI*. Boston, MA, USA: ACM, 2009, pp. 2751–2754.
- [21] M. Budwig, S. Jeong, and K. Kelkar, "When user experience met agile: a case study," in *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems CHI*. ACM, 2009, pp. 3075–3084.
- [22] J. T. Barksdale and D. S. McCrickard, "Concept mapping in agile usability: a case study," in *CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*. ACM, 2010, pp. 4691–4694.
- [23] J. Ferreira, H. Sharp, and H. Robinson, "Values and assumptions shaping agile development and user experience design in practice," in *XP 2010*, ser. Lecture Notes in Business Information Processing, vol. 48. Springer, 2010, pp. 178–183.
- [24] K. Tzanidou and J. Ferreira, "Design and development in the "agile room": Trialing scrum at a digital agency," in *XP 2010*, ser. Lecture Notes in Business Information Processing, vol. 48. Springer, 2010, pp. 372–378.
- [25] J. Ding, C. Lin, and K. Huang, "ARS: An adaptive reception scheme for handheld devices supporting mobile video streaming services," in *International Conference on Consumer Electronics. ICCE '06*, vol. 1, 2006, pp. 141–142.
- [26] Z. Hussain, M. Lechner, H. Milchrahm, S. Shahzad, W. Slany, M. Umgeher, T. Vlk, and P. Wolkerstorfer, "User interface design for a mobile multimedia application: An iterative approach," in *ACHI 2008, First International Conference on Advances in Computer-Human Interaction, Sainte Luce, Martinique, France*. IEEE Computer Society, February 10–15, 2008, pp. 189–194.
- [27] A. Cooper, *The Inmates Are Running the Asylum*. Indianapolis, IN, USA: Macmillan Publishing Co., Inc., 1999.
- [28] J. Pruitt and T. Adlin, *The Persona Lifecycle: Keeping People in Mind Throughout Product Design (The Morgan Kaufmann Series in Interactive Technologies)*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2006.
- [29] L. Shyba and J. Tam, "Developing character personas and scenarios: vital steps in theatrical performance and hci goal-directed design," in *C&C '05: Proceedings of the 5th conference on Creativity & cognition*. New York, NY, USA: ACM, 2005, pp. 187–194.
- [30] J. E. Nieters, S. Ivaturi, and I. Ahmed, "Making personas memorable," in *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM, 2007, pp. 1817–1824.
- [31] M. Pifarré and O. Tomico, "Bipolar laddering (bla): a participatory subjective exploration method on user experience," in *DUX '07: Proceedings of the 2007 conference on Designbiolaring for User eXperiences*. New York, NY, USA: ACM, 2007, pp. 2–13.
- [32] T. J. Reynolds and J. C. Olson, Eds., *Understanding Consumer Decision Making: The Means-end Approach to Marketing and Advertising Strategy*. Mahwah, NJ: Lawrence Erlbaum, 2001.

- [33] J. Nielsen, *Usability Engineering*. Morgan Kaufmann, 1993.
- [34] T. Hollingsed and D. G. Novick, "Usability inspection methods after 15 years of research and practice," in *SIGDOC '07: Proceedings of the 25th annual ACM international conference on Design of communication*. New York, NY, USA: ACM, 2007, pp. 249–255.
- [35] R. Molich and J. Nielsen, "Improving a human-computer dialogue," *Commun. ACM*, vol. 33, no. 3, pp. 338–348, 1990.
- [36] C. Wharton, J. Rieman, C. Lewis, and P. Polson, "The cognitive walkthrough method: a practitioner's guide," pp. 105–140, 1994.
- [37] S. Balbo, J. Coutaz, and D. Salber, "Towards Automatic Evaluation of Multimodal User Interfaces," in *IUI '93: Proceedings of the 1st International Conference on Intelligent User Interfaces*. New York, NY, USA: ACM Press, 1993, pp. 201–208.
- [38] M. L. Hammontree, J. J. Hendrickson, and B. W. Hensley, "Integrated data capture and analysis tools for research and testing on graphical user interfaces," in *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 1992, pp. 431–432.
- [39] M. Y. Ivory and M. A. Hearst, "The state of the art in automated usability evaluation of user interfaces," EECS Department, University of California, Berkeley, Tech. Rep. UCB/CSD-00-1105, Jun 2000. [Online]. Available: <http://www.eecs.berkeley.edu/Pubs/TechRpts/2000/5557.html>
- [40] R. Atterer and A. Schmidt, "Tracking the interaction of users with ajax applications for usability testing," in *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2007, pp. 1347–1350.
- [41] J. I. Hong and J. A. Landay, "WebQuilt: A Framework for Capturing and Visualizing the Web Experience," in *WWW '01: Proceedings of the 10th International Conference on World Wide Web*. New York, NY, USA: ACM Press, 2001, pp. 717–724.
- [42] M. Abdulkhair, "A multilingual automated web usability evaluation agent," Ph.D. dissertation, University of Sheffield, England, 2004.
- [43] M. Y. Ivory and M. A. Hearst, "Improving web site design," *IEEE Internet Computing*, vol. 6, no. 2, pp. 56–63, 2002.
- [44] M. Y. Ivory, "Web TANGO: Towards Automated Comparison of Information-centric Web Site Designs," in *CHI '00: CHI '00 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2000, pp. 329–330.
- [45] L. Paganelli and F. Paterno, "Automatic reconstruction of the underlying interaction design of web applications," in *SEKE '02: Proceedings of the 14th international conference on Software engineering and knowledge engineering*. New York, NY, USA: ACM, 2002, pp. 439–445.
- [46] F. Patern, A. Piruzza, and C. Santoro, "Remote web usability evaluation exploiting multimodal information on user behavior," in *CADUI*, G. Calvary, C. Pribeanu, G. Santucci, and J. Vanderdonckt, Eds. Springer, 2006, pp. 287–298.
- [47] S. J. Waterson, J. I. Hong, T. Sohn, J. A. Landay, J. Heer, and T. Matthews, "What Did They Do? Understanding Click-streams with the WebQuilt Visualization System," in *AVI '02: Proceedings of the Working Conference on Advanced Visual Interfaces*. New York, NY, USA: ACM, 2002, pp. 94–102.
- [48] Z. Hussain and W. Slany, "Analyzing real mobile web usage of a multimedia streaming application through log files," in *MIR '10: Proceedings of the international conference on Multimedia information retrieval*. ACM, 2010, pp. 401–404.
- [49] <http://www.agilemanifesto.org/> (last accessed 2011/12/01).
- [50] <http://zing.ncsl.nist.gov/WebTools/> (last accessed 2011/12/01).
- [51] <http://wauter.weeweb.com.au/> (last accessed 2011/12/01).
- [52] <http://www.goodgaze.com/ggx/> (last accessed 2011/12/01).
- [53] <http://www.feng-gui.com/> (last accessed 2011/12/01).
- [54] <http://www.usablenet.com/> (last accessed 2011/12/01).

# Emotion Recognition Through ANS Responses Evoked by Negative Emotions

## Emotion Recognition based on Machine Learning Algorithms

Eun-Hye Jang, Byoung-Jun Park, Sang-Hyeob Kim, Chul Huh

BT Convergence Technology Research Department  
Electronics and Telecommunications Research Institute  
Daejeon, S. Korea  
cleta4u@etri.re.kr; bj\_park@etri.re.kr;  
shk1028@etri.re.kr; chuh@etri.re.kr

Yeongji Eum, Jin-Hun Sohn

Department of Psychology/Brain Research Institute  
Chungnam National University  
Daejeon, S. Korea  
Yeongji\_Eum@gmail.com; jhsohn@cnu.ac.kr

**Abstract**—Emotion recognition using physiological responses is one of the core processes to implement emotional intelligence in human-computer interaction (HCI) research. The purpose of this study was to investigate emotion-specific ANS responses and test recognition rate using classification algorithm when negative emotion such as fear, surprise, and stress was evoked. The results of one-way ANOVA toward each parameter, there were significant differences among three emotions in skin conductance response (SCR), number of SCR (NSCR), skin temperature (SKT), and high frequency of HRV (HF). Results of emotion recognition applied to statistical method, i.e. linear discriminant analysis (LDA) and 4 machine learning algorithm, i.e. classification and regression tree (CART), self organizing map (SOM), Naïve Bayes and support vector machine (SVM) for emotion recognition showed that an accuracy of emotion classification by SVM was the highest and by LDA was the lowest. This can be helpful to provide the basis for the emotion recognition technique in HCI as well as contribute to the standardization in emotion-specific ANS responses.

**Keywords**-emotion recognition; machine learning algorithm; ANS responses

### I. INTRODUCTION

Emotion plays an important role in contextual understanding of messages from others in speech or visual forms. In advanced human-machine interaction, for affective communication between user and computer, it has to consider how emotions can be recognized and expressed during human-computer interaction and emotion recognition is one of the key steps towards emotional intelligence. Emotion recognition is one of the core processes to implement emotional intelligence in human-computer interaction (HCI) research [1]. Particularly, in important HCI applications such as computer aided tutoring and learning, it is highly desirable (even mandatory) that the response of the computer takes into account the emotional or cognitive state of the human user [2]. Emotion recognition has been studied using facial expression, gesture, voice, and physiological signal [3-7]. 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, measurement of emotional responses by multi-channel physiological signals offer more information for emotion recognition, because physiological responses are related to emotional state [8].

Many previous studies on emotion have reported that there is correlation between basic emotions such happiness, sadness, anger, etc. and physiological responses [9-15]. Also, experimental studies have been performed to distinguish specific emotions by autonomic nervous system (ANS) response and reported the emotion-specific ANS responses [10]. For example, in research reviewed 134 studies about ANS activity [11], ANS responses related to anger are a modal response pattern of reciprocal sympathetic activation and increased respiratory activity, particularly faster breathing. Studies on sadness report a heterogeneous pattern of sympathetic-parasympathetic coactivation and ANS responses of fear point to broad sympathetic activation, including cardiac acceleration, increased myocardial contractility, vasoconstriction, and increased electrodermal activity. Recently, emotion recognition using physiological signals has been performed by various machine learning algorithms, e.g., Fisher Projection (FP), k-Nearest Neighbor algorithm (kNN), Linear Discriminant Function (LDF), Sequential Floating Forward Search (SFFS), and Support Vector Machine (SVM). Previous works conducted a recognition accuracy of over 80% on the average seems to be acceptable for realistic applications [3-7]. Picard, Vyzas & Healey [3] classified 8 kinds of emotions (neutral, anger, grief, sadness, platonic love, romantic love, joy, & respect) by using kNN and it was verified 81.3% of accuracy. Haag, Goronzy, Schaich and Williams [4] applied MLP to categorize dimensions of arousal and valence in each emotion, and then it was reported as 80% of average accuracy. Also, Calvo, Brown and Scheduling [14] reported 42% of accuracy by using SVM to differentiate 8 kinds of emotions (neutral, anger, grief, sadness, platonic love, romantic love, joy, & respect).

However, Wagner, Kim and Andre [1] mentioned that it can be clearly observed that the accuracy strongly depends on the data sets which were obtained in laboratory conditions. That is, the results were achieved for specific users in specific contexts and it is very difficult to label emotion classes in physiological signals (waveforms) without uncertainty. Therefore, it is needed to recognize emotion

using manifold data sets obtained from different contexts and select the most optimal algorithm for emotion recognition. And in order to apply more elaborate feedback in HCI, it is necessary to discriminate the similar and vague emotions which are hard to be sorted. However, past studies is not a lot about discrimination of similar emotions under positive emotion or negative emotion likewise.

In this paper, to improve the limitation that it is result in specific context, we used 10 different emotional stimuli to induce one emotion under the same context conditions. And we verified the specific ANS responses of each emotion when negative emotions such as fear, surprise, and stress were evoked. The reason for selection of negative emotion was that in theory of evolution, it plays an important role in adaptation of living and surviving the evolution of human [16]. In particular, fear is associated with defensive aggression and is a kind of attempt to escape and people in fear emotion would attack only if escape is impossible [17-19]. Main function of surprise is to interrupt ongoing action and orient people to a possibly significant event. It has one core appraisal—appraising something as novel and unexpected—although other appraisals can shift the subjective feeling of surprise or shift the emotion from surprise to another emotion [20]. Stress are “constraining force or influence,” “a physical, chemical, or emotional factor that causes bodily or mental tension and may be a factor in disease causation,” and “a state resulting from stress—especially one of bodily or mental tension resulting from factors that tend to alter existent equilibrium [21].” Finally, we identified the optimal algorithm being able to classify three negative emotions. For this, we used a statistical method, linear discriminant analysis (LDA) which is one of the linear models, and 4 different machine learning algorithms, i.e., classification and regression tree (CART) of decision tree model, self organizing map (SOM) of Neural Network, Naïve Bayes of probability model, and SVM of non-linear model, which are used the well-known emotion algorithms.

II. METHODS

A. Participants



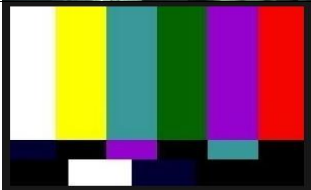
A total of 12 college students (6 males 20.8 years ± 1.26 and 6 females 21.2 years ± 2.70) participated in this study. They reported that they hadn’t had any history of medical illness or psychotropic medication and any kind of medication due to heart disease, respiration disorder, or central nervous system disorder. A written consent was obtained before the beginning of the study from the participants and they were also paid \$20 USD per session to compensate for their participation.

B. Emotional stimuli

Thirty emotional stimuli (3emotions x 10sets) which are the 2-4 min long audio-visual film clips captured originally from movies, documentary, and TV shows were used to successfully induce emotions (fear, surprise, and stress) in this study (TABLE 1). Fear-inducing films were the scene

which had tense and dreary atmosphere. Surprise clips were a section in which startling accident occurred and stress clip was TV adjustment scene that was mixture of black and white with white noise sound; you may easily see that when after all daily programs ended.

TABLE I. THE EXAMPLES OF EMOTIONAL STIMULI

Emotion	Contents	Examples
Fear	ghost, haunted house, scare, etc.	
Surprise	sudden or unexpected scream etc.	
Stress	audio/visual noise on screen, etc.	

The used audio-visual film clips were examined their suitability and effectiveness by preliminary study which 22 college students rated the category and intensity of their experienced emotion on emotional assessment scale after they were presented each film clip. The suitability of emotional stimuli means the consistency between the experimenter’s intended emotion and the participant’s experienced emotion (e.g., scared, surprise, and annoying). The effectiveness was determined by the intensity of emotions reported and rated by the participants on a 1 to 11 point Likert-type scale (e.g., 1 being “least surprising or least surprising or not surprising” and 11 being “most surprising”).

TABLE II. THE SUITABILITY AND EFFECTIVENESS OF EMOTIONAL STIMULI

	1	2	3	4	5	6	7	8	9	10	M
<b>Fe ar</b>	75 (10)	100 (9.9)	83 (9.8)	92 (9.6)	92 (9.7)	92 (9.7)	83 (9.6)	100 (9.3)	100 (9.3)	75 (8.7)	89 (9.6)
<b>Su rpr ise</b>	75 (9.3)	92 (9.7)	100 (9.7)	100 (9.9)	83 (9.6)	83 (9.6)	100 (9.5)	83 (9.4)	83 (8.6)	75 (10.3)	89 (9.5)
<b>Str ess</b>	92 (9.3)	100 (9.1)	100 (8.8)	100 (8.9)	100 (9.3)	100 (8.8)	92 (9.3)	100 (9.3)	100 (9.1)	100 (9.3)	98 (9.1)

above: suitability (%), below ( ): effectiveness (point)

The result showed that emotional stimuli had the suitability of 89% and the effectiveness of 9.1 point on average. The suitability of each stimulus was ranged from 75 to 100% and from 8.6 to 10.3 point in the effectiveness (TABLE 2).

C. Procedure

Prior to the experiment, participants were introduced to detail experiment procedure and had an adaptation time to feel comfortable in the laboratory’s environment. Then an experimenter attached electrodes on the participants’ wrist, finger, and ankle for measurement of physiological signals. Physiological signals were measured for 60 sec prior to the presentation of emotional stimulus (baseline) and for 2 to 4 min during the presentation of the stimulus (emotional state), then for 60 sec after presentation of the stimulus as recovery term. Participants rated the emotion that they experienced during presentation of the film clip on the emotion assessment scale (Figure 1). This procedure was repeated 3 times for elicitation of 3 differential emotions. Presentation of each film clip was count-balanced across each emotional stimulus. This experiment was progressed by the same procedures over 10 times.

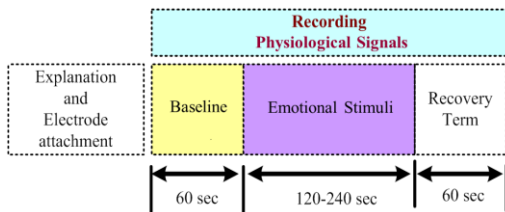


Figure 1. Experiment procedures

D. Experimental Settings

The laboratory was a sound-proof (lower than 35dB) room of 5m x 2.5m size where any outside noise or artifact was completely blocked. A comfortable chair was placed in the middle of the laboratory and 38 inch TV monitor for presentation of emotional stimuli was placed in front of the chair. An intercommunication device was located to the right side of chair for participant to communicate with an experimenter. A CCTV was installed on the top of the monitor set to observe participant’s behaviours and their behaviours were storage through the monitor and a video cassette recorder outside the laboratory.

E. Physiological measures and data analysis

The physiological signals were acquired by the MP100 system (Biopac System Inc., USA). The sampling rate of signals was fixed at 256 samples for all the channels. EDA was measured from two Ag/AgCl electrodes attached to the index and middle fingers of the non-dominant hand. ECG was measured from both wrists and one left ankle (reference) with the two-electrode method based on lead I. PPG and SKT were measured from the little finger and the ring finger

of the non-dominant hand, respectively. Appropriate amplification and band-pass filtering were performed.

The signals were acquired for 1minute long baseline state prior to presentation of emotional stimuli and 2-4 minutes long emotional states during presentation of the stimuli. The obtained signals were analyzed for 30 seconds from the baseline and the emotional state by AcqKnowledge (Ver. 3.8.1) software (USA) (Figure 2). The emotional states were determined by the result of participant’s self-report (scene that emotion is most strongly induced during presentation of each stimulus).

Features extracted from the physiological signals and were used to analysis are as follows: meanGSR, meanSCR, NSCR, meanSKT, maxSKT, meanPPG, Mean RR(s), STD(s), Mean HR(1/min), RMSSD(ms), NN50(count), pNN50(%), SD1(ms), SD2(ms), CSI, CVI, RR tri index, TINN(ms), FFTap\_LF, FFTap\_HF, ARap\_LF, ARap\_HF, FFTnLF, FFTnHF, FFTL/Hratio, ARnLF, ARnHF, and ARL/Hratio.

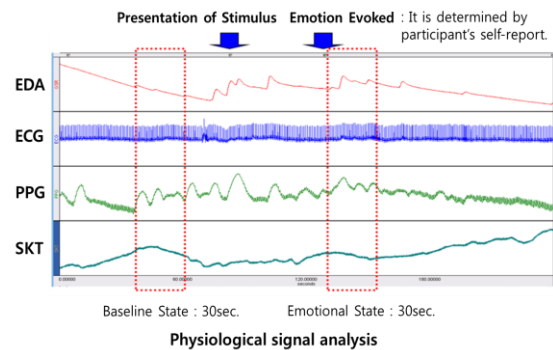


Figure 2. Analysis of physiological signals

270 physiological signal data (3 emotions x 10 stimuli x 9 cases) except for severe artifact effect by movements, noises, etc. were used for analysis. And differences of physiological signals among 3 emotions (alpha level at .05) were analyzed by one-way ANOVA (SPSS ver. 15.0). Also, to identify the emotion recognition algorithm being able to best recognize 3 different emotions by physiological signals, statistic method, LDA and 4 machine learning algorithms, i.e., CART which is a robust classification and regression tree, unsupervised SOM, Naïve Bayes classifier based on density, and SVM with the Gaussian radial basis function kernel were selected.

LDA, which is a statistical method to classify data signals by using linear discriminant functions, provides extremely fast evaluations of unknown inputs performed by distance calculations between a new sample and mean of training data samples in each class weighed by their covariance matrices [22]. CART is one of decision tree and nonparametric technique that 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 [23]. CART integrates the various information sources together for final decision. SOM, 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 [24]. The Naive Bayes algorithm is a classification algorithm based on Bayes rule and particularly suited when the dimensionality of the inputs is high [25]. SVM finds a hyperplane based on support vector to analyse data and recognize patterns. The complexity of the resulting classifier is characterized by the number of support vectors rather than the dimensionality of the transformed space [26]. Features' differences between emotional states and baseline extracted from physiological signals were used to apply these algorithms.

### III. RESULTS

#### A. Verification of the differences in ANS responses among emotions by one-way ANOVA

The results of one-way ANOVA using difference value of signals subtracting emotional states from baseline, there were statistically significant differences among three emotions in NSCR, mean SCR, mean SKT, max SKT and FFT ap\_HF (which is value to integrate an absolute value power of HF extracted from FFT) (TABLE III).

TABLE III. THE RESULT ON ONE-WAY ANOVA TOWARD EACH PARAMETERS

ANOVA	SS	df	MS	F
dNSCR	100.398	2	50.199	20.886***
dmeanSCR	7.363	2	3.681	6.242**
dmeanSKT	94.884	2	47.442	5.827**
dmaxSKT	91.563	2	45.781	5.744***
FFTap_HF	2,322.00	2	1,161.00	3.833*

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

To verify the difference among three emotions in detail, data were analyzed by LSD post hoc test. Figure 3 shows the result. There were significant differences of NSCR among all emotions and mean SCR between fear and stress, and between surprise and stress. SCR and NSCR, which are extracted from EDA decreased while all emotions were evoked, compared to baseline. Also, mean and max SKT distinguished between fear and surprise and between fear and stress. SKT decreased during fear induction and increased during surprise and stress from baseline. Finally, significant difference between fear and surprise was in FFT ap\_HF. There were an increase of FFT ap\_HF in fear and decreases of FFT ap\_HF in surprise and stress.

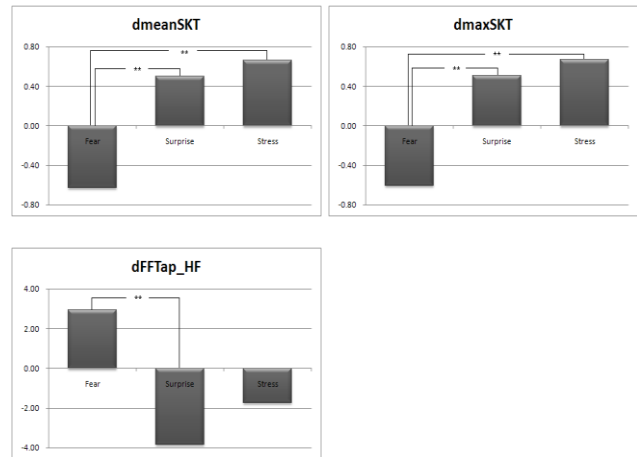
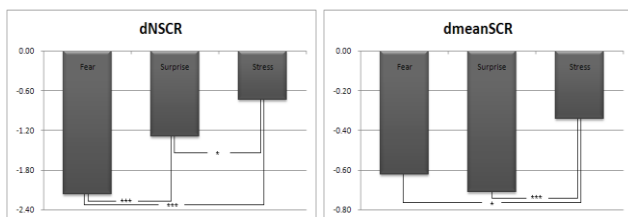


Figure 3. The results of LSD post-hoc test (\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ )

#### B. The results of emotion recognition by machine learning algorithm

28 features extracted from physiological signals were applied to 5 algorithms for emotion recognition of fear, surprise and stress. LDA, CART, SOM, Naïve Bayes and SVM were tested to confirm emotion recognition rate. The result of emotion recognition is like TABLE IV. 57.3% of originally grouped cases were correctly classified by LDA, 87.2% by CART, 59.5% by SOM, 80.9% by Naïve Bayes, and 100.0% by SVM. Three emotions, i.e., fear, surprise and stress were classified by SVM optimally.

TABLE IV. THE RESULT OF EMOTION RECOGNITION BY SVM

Algorithm	Accuracy (%)	Features (N)
LDA	57.3	28
CART	87.2	28
SOM	59.5	28
Naïve Bayes	80.9	28
SVM	100.0	28

The more detail results of emotion recognition accuracy by each algorithm are like from TABLE V to IX. In analysis of LDA, accuracy of each emotion had range of 49.5% to 62.0%. Fear was recognized by LDA with 60.6%, surprise 49.5%, and stress 62.0% (TABLE V).

TABLE V. THE RESULT OF EMOTION RECOGNITION BY LDA

	Fear	Surprise	Stress	Total
Fear	60.6	20.2	19.1	100.0
Surprise	22.2	49.5	28.3	100.0
Stress	16.0	22.0	62.0	100.0

CART provided accuracy of 87.2% when it classified all emotions and the recognition rate of each emotion was range of 84.0% to 92.2%. In fear, recognition rate of 92.2%

was achieved with CART, 85.4% in surprise, and 84.0% in stress (TABLE VI).

TABLE VI. THE RESULT OF EMOTION RECOGNITION BY CART

	<i>Fear</i>	<i>Surprise</i>	<i>Stress</i>	<i>Total</i>
Fear	92.2	5.0	3.0	100.0
Surprise	8.8	85.4	5.8	100.0
Stress	6.0	10.0	84.0	100.0

The result of emotion recognition using SOM showed that accuracy to recognize all emotions was 59.5%. According to orders of fear, surprise, and stress, recognition rates of 71.3%, 59.2%, and 48.0% were obtained by SOM (TABLE VII).

TABLE VII. THE RESULT OF EMOTION RECOGNITION BY SOM

	<i>Fear</i>	<i>Surprise</i>	<i>Stress</i>	<i>Total</i>
Fear	71.3	19.8	8.9	100.0
Surprise	26.2	59.2	14.6	100.0
Stress	30.0	22.0	48.0	100.0

The accuracy of Naïve Bayes algorithm to classify all emotion was 80.9%. And each emotion was recognized by Naïve Bayes with 83.2% of fear, 67.0% of surprise, and 93.0% of stress (TABLE VIII).

TABLE VIII. THE RESULT OF EMOTION RECOGNITION BY NAÏVE BAYES

	<i>Fear</i>	<i>Surprise</i>	<i>Stress</i>	<i>Total</i>
Fear	83.2	4.9	11.9	100.0
Surprise	15.5	67.0	17.5	100.0
Stress	4.0	3.0	93.0	100.0

Finally, accuracy of SVM was 100.0% and classifications of each emotion were 100.0% in all emotions (TABLE IX).

TABLE IX. THE RESULT OF EMOTION RECOGNITION BY SVM

	<i>Fear</i>	<i>Surprise</i>	<i>Stress</i>	<i>Total</i>
Fear	100.0	0.0	0.0	100.0
Surprise	0.0	100.0	0.0	100.0
Stress	0.0	0.0	100.0	100.0

#### IV. CONCLUSION

This study was to identify the difference among fear, surprise and stress emotions using physiological responses induced by these emotional stimuli and to find the optimal emotion recognition algorithm for classifying these three emotions. In our results, there were the differences of NSCR, mean SCR, mean SKT, max SKT, and FFT ap\_HF among emotions by one-way ANOVA. EDA index, i.e., NSCR and mean SCR, is signal that represents the activity of the

autonomic nervous system (activity of sweat glands) [27]. SKT variation reflects ANS activity and is effective indicator of emotional status. Variations in SKT mainly come from localized changes in blood flow, which is caused by vascular resistance or arterial blood pressure. The mechanism of arterial blood pressure variation can be described by a complicated model of cardiovascular regulation by the autonomic nervous system. Features of FFT ap\_HF extracted from ECG reflect the activity of cardiac activity. The sinoatrial (SA) node, which acts as pacemaker of cardiovascular activity, receives inputs from both branches (sympathetic and parasympathetic) of the autonomic nervous system. The activity level of the sympathetic nervous system is presented to the SA node by a postganglionic fibre, and that of the parasympathetic nervous system is given by a vagal nerve. The SA node can be thought of as a spike train generator whose inter-spike interval is modulated by the integration of the activity levels of the sympathetic and parasympathetic nervous system [12].

Our result showed that SVM is the best algorithm being able to classify fear, surprise and stress emotions. SVM is designed for two class classification by finding the optimal hyperplane where the expected classification error of test samples is minimized [28]. The SVM shows a recognition ratio much higher chance probability, i.e. 100.0% for three emotion categories, when applied to physiological signal databases. This was 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 were carefully extracted [12]. However, LDA and SOM had the lowest accuracy in emotion recognition. We think that this result in variability of physiological signals. The basic assumption that different emotions have a more or less unique and person-independent physiological response remains questionable. This could be reflected in the fact that the recognition rate falls off with the number of emotion categories [12]. These uncertainties could be an important cause that deteriorated the recognition ratio and troubled the model selection of the LDA or SOM.

Although some algorithm showed lower accuracy of emotion recognition, our results led to better chance to recognize human emotions and to identify the optimal emotion recognition algorithm by using physiological signals. In particular, SVM algorithm for classification of emotions can be helpful to provide the basis for the emotion recognition technique such as realization of elaborate and emotional man-machine interaction and will be applied to play an important role in several applications e.g., the human-friendly personal robot or other devices needed for various emotions recognition including how to develop effective robot control or adaptation behavior pattern using recognized emotion and product natural avatar's emotional behavior for interaction.

However, for more accurate and realistic applications, a novel method to identify not only basic emotions but also more various emotions such as boredom, frustration, and

love, etc. must be devised before it is mentioned that emotion recognition based on physiological signals is a practicable and reliable way of enabling HCI with emotion-understanding capability. Although, 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 [29]. Finally, 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. 2011K000655 and 2011K000658).

REFERENCES

[1] J. Wagner, J. Kim, and E. Andre, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, pp. 940-943, July 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] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 23, pp. 1175-1191, 2001.

[4] 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.

[5] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: First steps towards an automatic system," Affective Dialogue Systems, vol. 3068, pp. 36-48, 2004.

[6] J. A. Healey, "Wearable and automotive systems for affect recognition from physiology," Doctor of Philosophy, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[7] F. Nasoz, K. Alvarez, C. L. Lisetti, and N. Finkelstein, "Emotion recognition from physiological signals for user modeling of affect," International Journal of Cognition, Technology and Work-Special Issue on Presence, vol. 6, pp. 1-8, 2003.

[8] P. D. Drummond, and S. H. Quah, "The effect of expressing anger on cardiovascular reactivity and facial blood flow in Chinese and Caucasians," Psychophysiology, vol. 38, pp. 190-196, 2001.

[9] G. Stemmler, "The autonomic differentiation of emotions revisited: convergent and discriminant validation," Psychophysiology, vol. 26, pp. 617-632, 1989.

[10] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions," Science, vol. 221, p. 1208, 1983.

[11] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," Biological psychology, vol. 84, pp. 394-421, 2010.

[12] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," Medical & Biological Engineering & Computing, vol. 42, pp.419-427, 2004.

[13] J. N. Bailenson, E. D. Pontikakis, I.B. Mauss, J. Gross, M. E. Jabon, C. A. C. Hutcherson, C. Nass, and O. John, "Real-time classification of evoked emotions using facial feature tracking and physiological responses," International journal of human-computer studies, vol. 66, pp. 303-317, 2008.

[14] R. Calvo, I. Brown, and S. Scheduling, "Effect of experimental factors on the recognition of affective mental states through physiological measures," AI 2009: Advances in Artificial Intelligence, vol. 5866, pp. 62-70, 2009.

[15] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder," International journal of human-computer studies, vol. 66, pp. 662-677, 2008.

[16] C. Darwin, The expression of the emotions in man and animals. London: John Murray, 1872.

[17] D. C. Blanchard, and S. N. Takahashi, "Affect and aggression: An animal model applied to human behavior," Advances in the Study of Aggression, vol. 1, pp. 1-62, 1988.

[18] K. M. J. Lagerspetz, "Aggression and aggressiveness in laboratory mice," in Aggressive behavior, S. Garattini and E. B. Sigg, Eds., ed New York: Wiley, 1969, pp. 77-85.

[19] K. E. Moyer, The Psychobiology of aggression. New York: Harper & Row, 1976.

[20] K. R. Scherer, "Appraisal considered as a process of multilevel sequential checking," in Appraisal processes in emotion: Theory, methods, research, K. R. Scherer, et al., Eds., ed New York: Oxford University Press, 2001, pp. 92-120.

[21] Merriam-Webster. (2001, February 22). *WWWebster Dictionary [Online]*. Available: <http://www.m-w.com/dictionary.htm>

[22] M. Murugappan, R. Nagarajan, and S. Yaacob, "Classification of human emotion from EEG using discrete wavelet transform," Journal of Biomedical Science and Engineering, vol. 3, pp. 390-396, April 2010.

[23] J. R. Quinlan, C4.5 Programs for Machine Learning, San Mateo, CA: Morgan Kaufmann, 1992.

[24] F. Fdez-Riverola, E. L. Iglesias, F. Diaz, J. R. Mendez, and J. M. Corchado, "SpamHunting: an instance-based reasoning system for spam labeling and filtering," Decision Support Systems, vol. 43, pp. 722-736, 2007.

[25] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," Machine Learning, vol. 6, pp. 37-66, 1991.

[26] P. D. Wasserman, Advanced Methods in Neural Computing, New York, Van Nostrand Reinhold, pp. 35-55, 1993.

[27] W. Boucsein, "Electrodermal activity," Plenum Press, NewYork, 1992.

[28] 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.

[29] 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.

# Identification of Optimal Emotion Classifier with Feature Selections Using Physiological Signals

Byoung-Jun Park, Eun-Hye Jang, Sang-Hyeob

Kim, Chul Huh

IT Convergence Technology Research Laboratory  
Electronics and Telecommunications Research Institute  
Daejeon, South Korea

bj\_park@etri.re.kr, cleta4u@etri.re.kr,  
shk1028@etri.re.kr, and chuh@etri.re.kr

Jin-Hun Sohn

Department of Psychology & Brain Research Institute  
Chungnam National University  
Daejeon, South Korea  
jhsohn@cnu.ac.kr

**Abstract**—The purpose of this study is to identify optimal algorithm for emotion classification which classify seven different emotional states (happiness, sadness, anger, fear, disgust, surprise, and stress) using physiological features. Skin temperature, photoplethysmography, electrodermal activity and electrocardiogram are recorded and analyzed as physiological signals. For classification problems of the seven emotions, the design involves two main phases. At the first phase, Particle Swarm Optimization selects P % of patterns to be treated as prototypes of seven emotional categories. At the second phase, the PSO is instrumental in the formation of a core set of features that constitute a collection of the most meaningful and highly discriminative elements of the original feature space. The study offers a complete algorithmic framework and demonstrates the effectiveness of the approach for a collection of selected data sets.

**Keywords**—emotion classification, physiological signals, prototypes, feature selection, particle swarm optimization

## I. INTRODUCTION

Recently, the most popular research in the field of emotion recognitions is to recognize human's feeling using various physiological signals. In the physio-psychological research, it is known that there is strong correlation between human emotion state and physiological reaction. Psychologists and engineers have tried to analyze facial expressions, vocal emotions, gestures, and physiological signals in an attempt to understand and categorize emotions [1][2]. Most of all, physiological signals have been used to recognize human's emotions and feelings because the signals acquisition by non-invasive sensors is relatively simple, physiological responses induced by emotion are less sensitive in social and cultural difference, and there is a strong relationship between physiological reactions and emotional and affective states of humans [3].

Many previous studies on emotion have reported that there is correlation between basic emotions such as happiness, sadness, anger, fear, etc. and physiological responses [3][4]. Recently, emotion recognition using physiological signals has been performed by various machine learning algorithms, that is, Fisher's Linear Discriminant (FLD), k-Nearest Neighbor algorithm (kNN),

Support Vector Machine (SVM) and so on [5-8]. One of most widely applied machine learning method is instance-based learning (IBL [9]) which was shown to perform well in a number of challenging learning tasks.

In this paper, we introduce identification of optimal emotion classifier with feature selections using physiology signals induced by emotional stimuli for seven emotion classifications (happiness, sadness, anger, fear, disgust, surprise, and stress). To induce each emotion, ten emotional stimuli sets which have been tested for their suitability and effectiveness, are used in experiment. Physiological signals, namely, Skin temperature (SKT), photoplethysmography (PPG), electrodermal activity (EDA) and electrocardiogram (ECG) are acquired by MP100 Biopac system Inc. (USA) and analyzed to extract features for emotional pattern dataset. To complete their efficient development of a classifier for the seven emotions, we use one of the techniques of evolutionary optimization, namely, particle swarm optimization (PSO). In order to improve classification speed and classifier accuracy for the seven emotions, suitable formations of a set of prototypes and a core set of features are required. PSO embraces two level optimization processes. In the first level, PSO choose P % of patterns as a set of prototypes comes from patterns with seven emotional categories. In the second level of the optimization process, PSO is instrumental in the formation of a core set of features that is a collection of the most meaningful and discriminative components of the original feature space. Numerical experiments were carried out and it is shown that a suitable selection of prototypes and a substantial reduction of the feature space could be accomplished that is also accompanied with a higher classification accuracy using physiological signals for the seven emotions.

## II. MEASUREMENTS OF PHYSIOLOGICAL SIGNALS AND EMOTIONAL STIMULI

In this section, we will deal with experiments for the induction of the seven emotions and acquisition of physiological signals on an emotion induced by an emotional stimulus. 6 males (20.8 years±1.26) and 6 females (21.2 years±2.70) college students participated in this study. None of the subjects reported any history of medical illness or psychotropic medication and any medication that would

affect the cardiovascular, respiratory, or central nervous system. A written consent was obtained before the beginning of the experiment.

**A. Experimental Settings and Procedures**

The laboratory is a room of 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 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 behaviours and their behaviours were storage through the monitor and a video cassette recorder outside the laboratory as shown in Fig.1.

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 as shown in Fig. 2. Subjects rate the emotion that they experienced during presentation of the film clip on the emotion assessment scale. This procedure is conducted on each of the seven emotions for 10 times. 730 physiological signal data except for severe artefact effect by movements, noises, etc. are used for analysis.

**B. Emotional Stimuli and Physiological Signals**

To successfully induce the seven emotions (happiness, sadness, anger, fear, disgust, surprise and stress), seventy emotional stimuli, which consist of 10 sets for the seven emotions, are used in the experiments. Emotional stimuli are constituted 2~4 min long audio-visual film clips which are captured originally from movies, documentary and TV shows such as victory, wedding, laughing, etc. for happiness, death of parents/lover, separation, longing for mother, etc. for sadness, massacre, beating, attack, etc. for anger, ghost,

haunted house, scare, etc. for fear, body in pieces, vomiting, etc. for disgust, sudden or unexpected scream etc. for surprise, and audio/visual noise on screen, etc. for stress. 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 [10-12].

The suitability and effectiveness of emotional stimuli are examined in preliminary study prior to an experiment. The suitability of emotional stimuli means the consistency between the target emotions designed to induce each emotion and the categories of participants' experienced emotion. The effectiveness is determined by the intensity of emotions reported and rated by the subjects on a 1 to 11 point Likert-type scale (e.g., 1 being "least happy" or "not happy" and 11 being "most happy"). Twenty-two college students, that are different group from participants in the experiment, categorize their experienced emotion into seven emotion and estimate intensity of their categorized emotion on emotional assessment scale after being presented each film clip. The Table 1 shows the results of the suitability and effectiveness of emotional stimuli gotten from in preliminary study. The emotional stimuli have the suitability of 93% and the effectiveness of 9.5 point on average of 10 sets as shown in the results. The suitability of each stimulus is ranged from 75 to 100% and the effectiveness comes out from 8.4 to 10.4 point as shown in results.

The dataset of physiological signals such as skin temperature (SKT), electrodermal activity (EDA), photoplethysmography (PPG), and electrocardiogram (ECG) are collected by MP100 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.

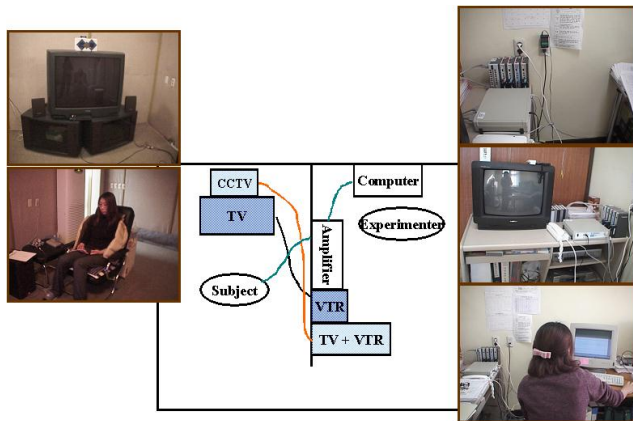


Figure 1. Example of Measuring Physiological Signals

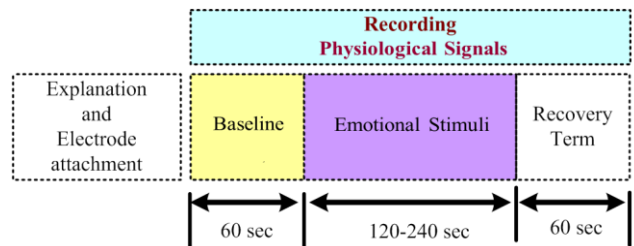


Figure 2. Experiment procedures

TABLE I. SUITABILITY AND EFFECTUALNESS OF EVOKING EMOTIONS

Set	Emotions	Happiness	Sadness	Anger	Fear	Disgust	Surprise	Stress	Average
1		100% (8.4)	92% (9.5)	75% (9.7)	75% (10)	75% (10.2)	75% (9.3)	92% (9.3)	83% (9.5)
2		100% (8.9)	100% (9.1)	75% (9.9)	100% (9.9)	92% (10.8)	92% (9.7)	100% (9.1)	94% (9.6)
3		100% (8.8)	100% (8.7)	75% (9.7)	83% (9.8)	92% (9.9)	100% (9.7)	100% (8.8)	93% (9.3)
4		100% (9.6)	100% (9.7)	75% (9.5)	92% (9.6)	100% (10.4)	100% (9.9)	100% (8.9)	95% (9.7)
5		100% (9.6)	100 % (9.3)	92% (9.8)	92% (9.7)	92% (9.7)	83% (9.6)	100% (9.3)	94% (9.6)
6		100% (9.3)	100% (9.3)	92% (9.4)	92% (9.7)	100% (10.3)	83% (9.6)	100% (8.8)	95% (9.5)
7		100% (9.3)	75% (8.9)	92% (8.9)	83% (9.6)	100% (9.3)	100% (9.5)	92% (9.3)	92% (9.3)
8		92% (8.0)	100% (9.0)	83% (9.2)	100% (9.3)	83% (10.2)	83% (9.4)	100% (9.3)	92% (9.2)
9		100% (9.7)	100% (9.2)	92% (9.5)	100% (9.3)	100% (10.1)	83% (8.6)	100% (9.1)	96% (9.4)
10		92% (8.8)	100% (9.3)	92% (9.7)	75% (8.7)	100% (10.1)	75% (10.3)	100% (9.3)	91% (9.5)
Average		98% (9.1)	96% (9.2)	84% (9.5)	89% (9.6)	98% (9.1)	89% (9.5)	98% (9.1)	93% (9.5)

The signals are acquired for 1 minute long baseline state prior to presentation of emotional stimuli and 2-4 minutes long emotional states during presentation of the stimuli. The obtained signals are analyzed for 30 seconds from the baseline and the emotional state by AcqKnowledge (Ver. 3.8.1) software (USA). The emotional states are determined by the result of participant’s self-report (scene that emotion is most strongly expressed during presentation of each stimulus). 28 features extracted from the physiological signals and used to analysis are as follows: SCL, NSCR, meanSCR, meanSKT, maxSKT, meanPPG, meanRRI [ms], stdRRI [ms], meanHR [1/min], RMSSD [ms], NN50 [count], pNN50 [%],SD1 [ms], SD2 [ms], CSI, CVI, RRtri, TINN [ms], FFTapLF, FFTapHF, FFTnLF, FFTnHF, FFTLF/HFratio, ARapLF, ARapHF, ARnLF, ARnHF, and ARLF/HFratio.

III. IDENTIFICATION OF OPTIMAL EMOTION CLASSIFIER

A. Two level processes for a classifier with feature selection

For the classification of the seven emotions, the proposed classifier is a type of instance based learning that uses only specific instances to solve classification problem. Namely, the classifier is a method for classifying objects based on the closest training patterns called prototypes in the feature space. This classifier embraces two selection problems to classify a new pattern to a class. One is the selection of prototype patterns and another one is feature selection.

In light of these observations, we adopt two level optimization processes for the formation of the prototypes and the feature space. To format prototype, we start with choose P % of patterns using particle swarm optimization (PSO). The classifier generates classification predictions using only P % of patterns. The classifier does not use any model to fit and only is based on distance between a pattern

and prototypes. Given a set of N prototypes, the classifier finds the one prototype closest in feature space to an unknown pattern, and then assigns the unknown pattern to the class label of its nearest prototype. The underlying distance between a pattern and a prototype is measured by weighted Euclidean one, that is

$$\|x - y\|^2 = \sum_{i=1}^n \frac{(x_i - y_i)^2}{\sigma_i^2} \tag{1}$$

Where **x** and **y** are the two patterns in the n-dimensional space and  $\sigma_i$  is the standard deviation of the i-th feature whose value is computed using the prototype set.

Secondly, once the prototypes have been formed, we reduce feature space by choosing a core set of features encountered in the problem. Those features are regarded as the most essential ensemble of features that, organized together, exhibit the highest discriminatory capabilities. Often their number could be quite limited in comparison with the dimensionality of the overall feature space. One can consider d % of the total number of features, say 10%, 20%, etc. The features forming the core have to be considered altogether. We use PSO to choose d % of features which minimizes the classification error.

Overall, the algorithm can be outlined as the following sequence of steps

Step 1: Randomly generate “N” particles,  $p_i$ , and their velocities  $v_i$ . Each particle in the initial swarm (population) is evaluated using the objective function. For each particle, set  $pbest_i = p_i$  and search the best particle of  $pbest$ . Set the best particle associated with the global best,  $gbest$ .

Step 2: Adjust the inertia weight, w. Typically, its values decrease linearly over the time of search. We start with  $w_{max} = 0.9$  at the beginning of the search and move down to  $w_{min} = 0.4$  at the end of the iterative process,

$$w(t) = w_{\max} - \frac{w_{\max} - w_{\min}}{\text{iter}_{\max}} \times t \tag{2}$$

where  $\text{iter}_{\max}$  denotes the maximum number of iterations of the search and “t” stands for the current index of the iteration.

Step 3: Given the current values of **gbest** and **pbest<sub>i</sub>**, the velocity of the i-th particle is adjusted following (3). If required, we clip the values making sure that they are positioned within the required region.

$$v_i = wv_i + c_1r_1(\text{pbest}_i - p_i) + c_2r_2(\text{gbest} - p_i) \tag{3}$$

Step 4: Based on the updated velocities, each particle changes its position using the expression

$$P_{ik} = v_{ik} + P_{ik} \tag{4}$$

Step 5: Move the particles in the search space and evaluate their fitness both in terms of **pbest<sub>i</sub>** and **gbest**.

Step 6: Repeat from Step 2 to Step 5 until the termination criterion has not been met. Otherwise return **gbest** as the solution found.

**B. Prototypes and features versus PSO**

The underlying principle of PSO[13] involves a population-based search in which individuals representing possible solutions carry out a collective search by exchanging their individual findings while taking into consideration their own experience and evaluating their own performance. PSO involves two competing search strategy aspects [14][15]. One deals with a social facet of the search; according to this, individuals ignore their own experience and adjust their behavior according to the successful beliefs of individuals occurring in their neighborhood. The cognition aspect of the search underlines the importance of the individual experience where the element of population is focused on its own history of performance and makes adjustments accordingly. PSO is conceptually simple, easy to implement, and computationally efficient. Unlike many other heuristic techniques, PSO has a flexible and well-balanced mechanism to enhance the global and local exploration abilities [15]. The basic elements of PSO technique are performance index (fitness), particles, best particles, and velocity. The particle is moving in the search space with some velocity which plays a pivotal role in the search process.

As PSO is an iterative search strategy, we proceed until

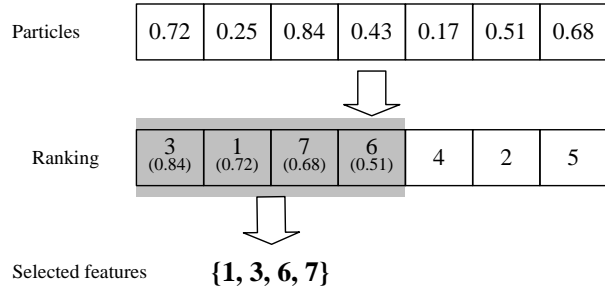


Figure 3. PSO formation of the reduced feature space; here, n=7 while the assumed reduction of the space results in 4 features viz. {1, 3, 6, 7}.

there is no substantial improvement of the fitness or we have exhausted the number of iterations allowed in this search. As a generic search strategy, the PSO has to be adjusted to solve a given optimization problem. There are two fundamental components that deserve our attention: a fitness function, and a representational form of the search space. Given the prototype formation and feature selection, we consider the minimization of the classification error as a suitable fitness measure. There could be different choices of the search space considering that an optimal collection of features could be represented in several different ways. Here we adopt the representation scheme of the search space in the form of the n-dimensional unit hypercube. The content of a chromosome is ranked viz. each value in this vector is associated with an index the given value assumes in the ordered sequence of all values encountered in the vector. Considering that we are concerned with d % of all features, we pick up the first d×n (0<d<1) entries of the vector of the search space. This produces a collection of features forming the reduced feature space. This mechanism of the formation of the feature space is portrayed in Fig. 3.

For the entire patterns, the prototype formation is carried out in the same manner as we encountered in the feature selection.

**IV. NUMERICAL EXPERIMENTS**

The numerical studies presented here provide some experimental evidence behind the effectiveness of the PSO approach. The detailed setup of an extensive suite of experiments is reflective of the methodology we outlined in the previous sections. The two essential parameters that we use in the assessment of the performance of prototype and feature selection are the percentage of features (denoted by

TABLE II. CLASSIFICATION ACCURACY (AVERAGE AVG AND STANDARD DEVIATION STD %) REGARDED AS A FUNCTION OF “d” AND “P” FOR THE SEVEN EMOTIONAL DATASET

d % (No. of F)	P %			AVG±STD over P
	30	50	70	
30 (8)	58.3 ± 3.89	79.5 ± 10.93	90.3 ± 6.36	76 ± 15.42
50 (14)	45.5 ± 4.15	54.1 ± 5.04	72.4 ± 4.39	57.3 ± 12.24
70 (20)	35.5 ± 1.92	39.9 ± 3.41	47.3 ± 3.74	40.9 ± 5.81
100 (28)	35.5 ± 1.86	39.4 ± 1.9	49.9 ± 2.15	41.6 ± 6.46

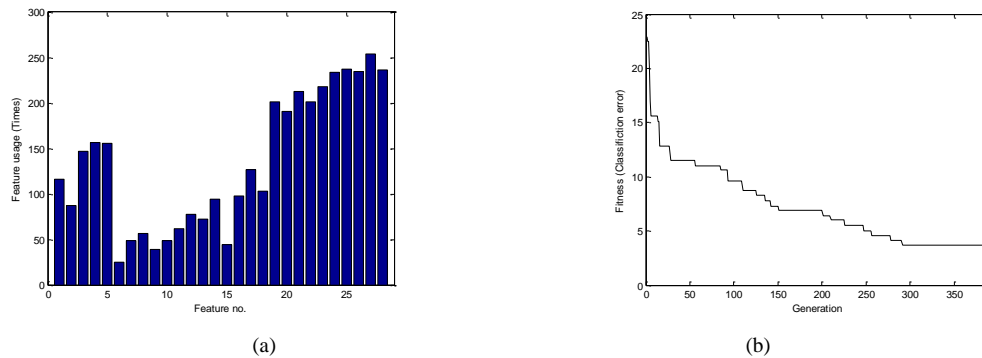


Figure 4. Results of feature selections on PSO: (a) feature usage index over all values of d and P, and (b) a fitness function of PSO for d=30 and P=70

“d”) forming the core of the reduced feature space and the percentage of the data forming the prototype set (P) optimized by the PSO. The results are reported for the testing data sets for various values of “P” and “d”.

The numeric values of the parameters of the PSO were either predetermined (considering some existing guidelines available in the literature, cf. [16][17]) or it becomes selected experimentally. More specifically, we used the following values of the parameters: maximum number of generations is 500; swarm size is 150; maximal velocity,  $v_{max}$ , is 20% of the range of the corresponding variables;  $w_{min}=0.4$ ;  $w_{max}=0.9$ ; and acceleration constants  $c_1$  and  $c_2$  are set to 2.0. The inertia weight factor “w” has been regarded as a linearly decreasing function of iterations (optimization time)

The seven emotional classes consist of happiness, sadness, anger, fear, disgust, surprise and stress. The proposed prototype-based classifier is evaluated by 10 times repeated random sub-sampling validation for seven emotion classification. The 70% of the whole emotional patterns are selected randomly for training and the remaining patterns are used for testing purposes. The results are reported by presenting the average and standard deviation of the classification accuracy obtained over 10 repetitions of the experiment for the test dataset. When reporting results, we

concentrate on the determination of relationships between the collections of features and obtained classification rates. We also look at the optimal subsets of features constructed with the use of the method.

For the seven emotions dataset given as multi-physiological signals, the relationship between the percentage of features used in the PSO optimization, values of “P” and the resulting classification accuracy is presented in Table II. Here, “No. of F” is the number of selected features for d % of entire features, “AVG” and “STD” indicate average and standard deviation, respectively.

The classification accuracy was computed over 10-fold realization of the experiments, namely, for each combination of the values of the parameters (d and P), the experiments was repeated 10 times by running PSO.

With the increasing values of “d”, the classification accuracy of the seven emotions decreases substantially; in the case of P=30% it drops from 58.3 to 35.5 when increasing the number of features from 30% to 70%. The similar downward tendency occurs when dealing with any P % and considering the same increase in the percentage of features. Conclusively, the use of all features dropped accuracy of classification for the seven emotions. Changes in the values of “P” have far less effect on the classification rate, however, the distinguished result was occurred in d=30 and P=70. From these results, we observe that the number of suitable features is 8 and 70 % of dataset are required as the prototype for the seven emotion recognition using multi-physiological signals. We report the number of occurrences of the features in Fig. 4 (a). The number of occurrences of a given feature is computed across all values of “P” and “d”. Interestingly, there are several dominant features such as FFTnHF (feature24), FFTLF/HFratio (feature 25), ARnLF (feature 26), ARnHF (feature 27), and ARLF/HFratio (feature 28). meanPPG (feature 6), meanHR (feature 9) and CSI (feature 15) are of lowest relevance. Prototypes were picked up by PSO as a supervisor of the classifier for each class and given patterns were assigned into a class through those and selected features. In case of d=30 for the classification of the seven emotions, namely, the number of features is 8, we have gotten that classification accuracy is 90.3 % for P=70. PSO Fig. 4 (b) shows the fitness function of PSO. Therefore, we can consider than the number of core

TABLE III. COMPARISON OF THE CLASSIFICATION ACCURACY OF THE PROPOSED METHOD AND OTHER METHODS (RESULTS FOR TESTING DATA)

Method	Accuracy (%)	Number of Features
CART	21.7 ± 3.6	28
C4.5	16.1 ± 1.2	28
kNN	45.3 ± 2.3	28
FLD	20.3 ± 1.8	28
NN	18.0 ± 1.0	28
PNN	16.3 ± 1.9	28
RBFs	17.4 ± 1.2	28
SOM	Supervised	16.3 ± 1.9
	Unsupervised	15.4 ± 3.2
SVM	17.8 ± 3.1	28
Proposed Methodology	90.3 ± 2.26	8



features is eight and feature 19, 21, 23, 24, 25, 26, 27, and 28 are selected on Fig. 4 (a).

For the classification of the seven emotions, Table III contrasts the classification accuracy (%) of the proposed method with other well-known methods studied in the literatures [5-8]. As abovementioned, the experiments are reported for the 10 times using a split of data into 70%-30% training and testing subsets, namely, 70% of the whole patterns are selected randomly for training of all methods and the remaining patterns are used for testing purposes. The results are averaged over 10 times for testing dataset. The experimental results reveal that the proposed approach and the resulting model outperform the existing methods both in terms of the simpler structure and better prediction (generalization) capabilities on feature space reduced 70% of entire feature space.

#### V. CONCLUSION

In this study, we have discussed the acquisition of multi-physiology signals using emotion stimuli and the design of a classification methodology for the seven 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 suitability of 93% and the effectiveness of 9.5 point on average. In addition that, we have introduced an instance-based learning classifier with feature selection learned by particle swarm optimization (PSO) mechanism for the seven emotions expressed by multi-physiological signals. The optimization process of forming the prototypes and the feature space is reflective of the conjecture on the importance of forming a set of prototypes and a core set of features whose discriminatory capabilities emerge through their co-occurrence in these set. The methodology of feature selection becomes legitimate considering that we immediately see the result of the reduction of the feature space being translated into the corresponding classification rate. The use of the prototype is also justifiable considering that this classification scheme is the simplest that could be envisioned in pattern classification.

#### ACKNOWLEDGMENT

This research was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (No. 2011K000655 and 2011K000658).

#### REFERENCES

- [1] 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.
- [2] F. Nasoz, K. Alvarez, C. L. Lisetti, N. Finkelstein, "Emotion recognition from physiological signals for presence technologies," *Cognition, Technology and Work, Special Issue on Presence*, Vol. 6, pp.4-14, 2004.
- [3] O. Alaoui-Ismaili, O. Robin, H. Rada, A. Dittmar and E. Vernet-Maury, "Basic emotions evoked by odorants: comparison between autonomic responses and self-evaluation," *Physiology and Behavior*, vol. 62, pp. 713-720, 1997.
- [4] 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.
- [5] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and regression trees*. Monterey, Calif., U.S.A.: Wadsworth, Inc.
- [6] J.R. Quinlan, *C4.5 Programs for Machine Learning*, San Mateo, CA: Morgan Kaufmann, 1992.
- [7] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern classification*, 2nd edn. Wiley-Interscience, New York, 2000.
- [8] T. Kohonen, *Self-Organizing Maps*, Springer Series in Information Sciences, Vol. 30, Springer, Berlin, Heidelberg, New York, 1995, 1997, 2001. Third Extended Edition, pp.501.
- [9] D.W. Aha, D. Kibler, and M.K. Albert, "Instance-based learning algorithms, *Machine Learning*," vol. 6, pp. 37-66, 1991.
- [10] 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.
- [11] J. J. Gross, and R. W. Levenson, "Emotion elicitation using films," *Cognition and Emotion*, vol. 9, pp. 87-108, 1995.
- [12] 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.
- [13] J. Kennedy, "The particle swarm: social adaptation of knowledge," in *Proc. IEEE Int. Conf. Evolutionary Comput.*, 1997, pp. 303-308.
- [14] C.-L. Huang and C.-J. Wang, "A GA-based feature selection and parameters optimization for support vector machines," *Expert Systems with Applications*, vol. 31, pp. 231-240, 2006.
- [15] M.A. Abido, "Optimal design of power-system stabilizers using particle swarm optimization," *IEEE Trans. Energy Conversion*, vol. 17, pp. 406-413, 2002.
- [16] Z.-L. Gaing, "A particle swarm optimization approach for optimum design of PID controller in AVR system," *IEEE Trans. Energy Conversion*, vol. 19, pp. 384-391, 2004.
- [17] R. Eberhart and Y. Shi, "Particle swarm optimization: developments, applications and resources," in *Proc. IEEE Int. Conf. Evolutionary Comput.*, 2001, pp. 81-86.

# Interacting with Navigation Devices: A Case Study

Javier Calle, Esperanza Albacete, Enrique Sánchez, Dolores Cuadra  
 Computer Science Department – Carlos III University of Madrid  
 Avd. Universidad, 30, 28911 (Spain)  
 {fcalle, ealbacet, essotes, dcuadra}@inf.uc3m.es

**Abstract**— The paper aims to study the adequacy of interactive capabilities in commercial navigation devices. In order to do this, an experiment is proposed and two sets of results are obtained and analyzed: one resulting from the application of objective metrics; and the other one, from subjective answers of the device users when queried about their experience. To the extent that these interactive capabilities are found to be inadequate, it is the goal of this study to ascertain (1) whether a more natural, human-like interaction paradigm (i.e., natural interaction) ought to be incorporated into these kinds of devices and, if so, (2) the extent to which their interactive capabilities improve as a result. For this, a direct comparison of device interactive capabilities with those of human beings is necessary. Study results suggest that the incorporation of natural interaction in navigation devices may help respond to certain device interactive inadequacies as well as to the user dissatisfaction which those inadequacies engender.

*Keywords*- Guiding Service; Natural Interaction; Evaluation

## I. INTRODUCTION

Among the important challenges in the technological field today is that of the selection of the most suitable mechanisms by which particular devices interact with. It is broadly accepted that, for a given device, the most suitable interface is not necessarily the most complex one. Some interfaces have been used over many years and thus enjoy a high level of user familiarity (e.g., it is hard to imagine how to improve traffic lights for the sharing of a crossroads). In fact, for most devices, the simplest interface is often thought to be the best. For this reason, current trends point to minimalist interfaces [13], even for complex devices. Given particular contextual information, then, these devices should be able to find the right command. However, their use is not necessarily simple and may require including additional interfaces, leaving the minimalist interface as a shortcut.

With users specifically in mind, intuitive interfaces are often sought out that are icon-based and supported by touch screens. However, as discussed in other studies [5], the ultimate user-oriented goal is achieving a system which behaves in a more human-like way and, more specifically, interacts with users in the same way that those users might do with other human beings. For this reason, voice technologies are applied to many systems – since they are more familiar and accessible for most users – and often complemented with a graphical user interface (GUI). Additionally, several systems have also been developed to emulate other aspects of human interaction such as dialogue strategies, turn-taking, adaptation to the interlocutor and contextual awareness.

The present paper explores the suitability of this sort of interaction paradigm in navigation devices. Since the tasks for which these devices are created are also regularly performed collaboratively by human beings, one may think that the application of the NI paradigm should be highly beneficial. Not surprisingly, then, recent years have seen the development of prototypes of multimodal interaction systems for guiding by several research groups in the field [10][15]. However, before such costly research is undertaken, it is advisable first to conclusively determine (1) whether the NI paradigm is, in fact, suitable for incorporation in navigation devices and (2) whether current navigation devices already possess sufficiently adequate interactive capabilities (i.e., without the incorporation of the NI paradigm).

In the present case study, the interactive capabilities of a commercial device for pedestrian navigation supported by GPS are studied. In order to observe the results obtained through the incorporation of the NI paradigm in the device, a human tester is used to simulate the role of a more human-like interface, while not interfering with the tasks performed.

Finally, it is important to remember that collaborative task performance does not necessarily mean that participants are exclusively committed to its execution. Quite to the contrary, participants often juggle several concurrent goals—some shared and others individual. With this fact in mind and in order to ensure the realism in the experiments, both participants are assigned a second, background collaborative task to be performed simultaneously with the first.

## II. RELATED WORK ON HUMAN-LIKE INTERACTION

Primarily for the purpose of achieving greater accessibility, NI research seeks to endow systems with more natural interactive capabilities. In this way, users with relatively little to no technological training are able to access information, applications and services provided by the system [5]. Furthermore, this sort of interaction offers a convenient and appealing alternative for more technologically literate individuals, as well, particularly when traditional modes of interaction become cumbersome.

One of the principal features of human interaction is its flexibility with regard to both the conceptualization and expression of utterances. Human interlocutors exploit a wide range of lexical and expressional resources and dialogue strategies at their disposal. For quick adaptation to changes in the interaction, particular resources and strategies utilized by the interlocutor may be substituted for others at any moment. The reasons for this interactive behavior can be understood according to Grice's conversational Maxims [8]:

be sufficiently but not excessively informative (Maxim of Quantity), be truthful (Maxim of Quality), be relevant (Maxim of Relation) and be clear (Maxim of Manner). These Maxims have already been used by Harris [9] to support the evaluation of voice systems.

It is important to emphasize, however, the fluid and overarching quality of the Gricean Maxims. In order to obtain a particular (and often non-literal) conversational result, for instance, Maxims may be consciously flouted by an interlocutor [3], essentially converting the Maxims into extensions of the dialogue strategies in the interlocutor's arsenal. Additionally, interaction and its interpretation according to the Gricean Maxims in a particular case, is framed by different contextual aspects which may be classified as material (i.e., time, place, weather ...), semiotic, political, operative and socio-cultural environment [7].

Other important elements frequent in human interaction are interruptions [1]. Individuals are almost always in the process of performing tasks to achieve specific goals. While they may collaborate in the execution of a shared task, it is most often the case that those task participants are also simultaneously involved in the execution of several other tasks. When a particular individual needs to communicate information, this act will almost always result in the interruption of another task being executed by the receiver [12]. Even after the participants begin interacting, an utterance by one of them to the other may nevertheless be considered an interruption if either the former or latter participant happened, at the same time, to be in the middle of a different utterance (i.e., *self-interruptions and locutive interruptions*, respectively), to be talking about a different topic (i.e., *dialogue line interruption*) or to be using a different dialogue strategy (i.e., *development interruption*). A participant desiring make such an interruption decides to act or to refrain from acting on this desire based on an analysis of the costs and benefits resulting from the interruption [11][14]. However, those costs may nevertheless be reduced either through the employment of certain interaction techniques, such as pre-sequencing and displays of awareness [6], or by choosing the proper moment [2].

To endow a system with these interactive capabilities and characteristics, NI research seeks to gather knowledge about human interaction and reasoning mechanisms. A main trend in NI research is the classification and distribution of knowledge involved in the human interaction process across specialized models. Among some of the different types of models proposed in a previous study [4], interaction structures and intentions would be handled by a dialogue model, turn-taking by a presentation model, interlocutor characterization by a user model and context management by a situation model. Other frequently included components are the emotional models, task models, system (or self) models, ontology and, of course, interface components. Each particular knowledge distribution depicts an NI cognitive architecture, several examples of which being found in the literature [16]. These models are often implemented by autonomous agents [5] continuously processing the interaction and cooperating to achieve the global goal of a more natural human-like interaction.

### III. CASE STUDY EXPERIMENTS

For the sake of clarity, the section is divided into five sub-sections detailing the design, preparation and results of the experiments. The discussion can be found in Section 4.

#### A. Experiment Design

The experiments conducted for this article studied the interaction between two individuals playing the roles of user and guide, where the former is the test subject and the latter the experiment leader. In addition to the latter's role as guide, the experiment leader was also responsible for recording the interactions with the test subject, the breaks taken between experiments and the later interview of the test subject following the completion of the experiments. In addition to these two participants, a third participant shadowed the former two as an objective observer, manually noting down any important events that arose during the dialogue without ever directly intervening in the interaction itself. This observer was also responsible for conducting the post-experimental interview of the test subject.

The interaction studied in these experiments was developed around the simultaneous execution of two distinct, high-level tasks: (1) a principal task of guiding a user along a route from a fixed start point to a fixed end point and (2) a background task of exchanging information (i.e., chatting) about a set of films proposed by the test subject. The background task was incorporated into the experiments in order to produce real human interactions in which participants are often engaged in the simultaneous execution of multiple tasks. Furthermore, the specific topic of this background task was chosen due to its emotionally-engaging nature, thereby commanding higher interactive commitment from the user [4]. Consequently, the importance given by the user to the main task, while still high, would nevertheless not be absolute and any unnecessary use of resources in its development, therefore, would detract from the development of the emotionally-engaging background task and most likely annoy the user. Finally, walking routes are preferred since they reduce risks and make easier to observe the test subject. Besides, and insofar as the experiments took place in a real environment requiring participant attention (e.g., awareness of road conditions to avoid being hit by a car), a third, low-level task of environmental awareness was also considered.

Consistent with what has been discussed in previous sections of the article, the main goal of these experiments was to observe the oral interactions produced by navigation devices (GPS) and to contrast them with human interactions produced in the same operational context. The experiments were supported by a GPS device chosen from among the most popular, commercially available models. It was carried by the experiment leader who hid GPS visual displays while ensuring that audio utterances were perfectly audible for the user. The leader was allowed to provide additional information for the development of the main, high-level task, but only if absolutely necessary and strictly limited to verbal utterances (i.e., avoiding spatial deixis through gestures).

Since events occurring in the test environment during the execution of the principal high-level task and the low-level task were likely to grab the participants' attention, several

different navigation routes were developed and executed in the experiments by each test subject. The specific routes were designed for the experiments to reflect different levels of navigational difficulty through varied route lengths, numbers of hotspots (i.e., points where the user is more likely to get lost) and environmental elements (i.e., elements requiring different levels of attention to the low-level task).

The routes were set prior to the execution of the experiments and were the same for all test subjects. Three routes were developed: one comprised entirely of roads, another comprised entirely of walkways and the third comprised of both roads and walkways. The routes were located to be executed in succession while, at the same time, allowing for a short break in a quiet area between each task. The routes were designed to take an hour to cover them all (including breaks). The test subjects were screened prior to the experiments to ensure that none was familiar with the destinations of the different routes or even the majority of roads or walkways traversed in order to reach those destinations. The experiment crew formed by the experiment leader and observer received training on the routes to be followed prior to the execution of the experiments.

*B. Participant Selection and Experiment Preparation*

In a preparatory phase of the experiments, the specific GPS device to be used, the routes to be followed and the test subjects to participate as users in the study were all preselected. For the selection of the device, a Garmin nüvi@ navigator was chosen due to its wide commercial use and popularity. The three routes for the experiments were developed in the Spanish city of Leganés due its proximity to Madrid and fulfillment of the environmental requirements discussed earlier. The total length of the routes developed was less than 3 kilometers with an estimated total completion time – including breaks and travel times between routes – of less than one hour. Nevertheless, it was understood that actual route lengths and estimated completion times would likely vary from subject to subject, particularly in the case that a test subject misunderstood the instructions or selected a path different from that given by the navigator. Specific details about these routes (i.e., road, walkway and hybrid routes) are presented below in TABLE I.

TABLE I. DESCRIPTION OF ROUTES DEVELOPED

Route	Roads (#)	Walkways (#)	Hotspots (#)	Length (m)	Time (min)
Hybrid	6	3	5	1100	20
Roads	4	-	7	850	14
Walkways	-	3	4	500	7

As noted by the observer, road routes in the experiments appeared easier for navigation than others due to the restricted space for pedestrian transit and the fewer points along the route (i.e., hotspots) requiring a navigational decision by the test subject. However, the traffic likely to be present in the test environment would nevertheless pose greater dangers for the test subject, therefore requiring a greater dedication of interaction resources to the execution of the low-level task (i.e., environmental awareness).

Regarding the selection of test subjects, 50 candidates were interviewed with the aim of obtaining a wide distribution of primary characteristics – such as age and experience with GPS devices – and secondary characteristics – such as personal sense of direction, experience with computers and experience with technological devices, in general. Of the candidates interviewed, 22 were selected in such a way as to maximize these distributions of primary and secondary characteristics, thereby producing more generalizable results. In order to reduce any potential effects of the route sequencing, all selected test subjects were given a training session prior to the experiments which included a description of the process and a brief trial run.

*C. Data Acquisition*

Experiments in this study were organized such that the data recorded would allow researchers to observe to what extent Gricean Maxims (see Section II) were met in the interaction between the GPS device and test subjects. While all Gricean Maxims were considered, less importance was given here to the Maxim of Quality insofar as a lack of “truthfulness” would not reflect the interaction quality, but rather the positioning technology precision, which is not the focus of the present study. Observations regarding temporal realization and interruptions were of particular interest, since both of them are characteristic of human-like interaction, yet frequently overlooked by classical interaction systems.

In order to allow for the natural execution of tasks, it was deemed most appropriate here to employ a procedure of external observation followed by the subsequent collection of test subject impressions regarding the interaction and the experiments themselves. According to this procedure, the observer noted all important details and events in each case with specific attention paid to the details discussed below.

Regarding the Gricean Maxim of Quantity, any moments where the device provided the user with either insufficient or excessive navigational information were to be noted down by the observer. They were respectively measured as user utterances made to request omitted information and useless interventions, both regarding total number of utterances.

Nevertheless, as redundancy is not necessarily a negative indicator, but may, in fact, be required in certain cases, redundant device utterances were noted down by the observer, classified as either useful or superfluous and later analyzed. Where a device utterance was considered to be superfluous, the observer was also to detail the user’s reaction to the utterance, noting if the user responded with any signs of annoyance. Besides, as relevant information may nevertheless fail to fully produce its desired effect if not framed in the proper manner, the experiment observer was charged with recording the number of times a user requested clarification, hesitated or made an incorrect decision.

All these observations were presented together according to the routes during which they were originally recorded. Since each route is of a different length and possesses a different number of hotspots, the effective comparison of results recorded requires a prior normalization of the data. For this normalization, the average execution time and number of decisions point for each route was calculated.

Apart from the details described above, the observer was instructed to record any comment or gesture made by the test subject during routes. Through breaks no notes were taken, for the test subject to feel less scrutinized and more willing to talk freely about the interaction experience. Comments made by the test subject during pauses were transcribed from the recordings following the completion of the experiments.

In order to complete the subjective data obtained during the experiments, then, open interviews of test subjects were conducted (following the completion of the experiments) in which the interviewer (i.e., the observer) focused on five topics of specific relevance to the aims of the study: (1) the quantity of information provided, (2) the usefulness of the information provided, (3) the clarity of device utterances, (4) the timing of the utterances, and (5) the subjects' preference to interact with a human being or a device in the execution of navigational tasks. Just as occurred during the pauses between routes and in order to ensure a relaxed interview environment and lively test subject responses, no manual notes were taken during the interview. Instead, questions and responses were recorded electronically and later transcribed.

**D. Objective Results from Observation**

Following the completion of the experiments, the data collected by the observer was prepared for analysis. Most interesting results recorded are presented in the following graphs. In Figure 1.a, the number of regular, redundant and clarificatory device utterances per hour is shown. The data is presented both for each specific route studied as well as for the experiments as a whole (i.e., the totality of the routes).

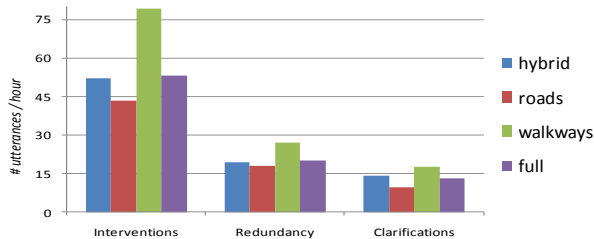


Figure 1.a: System utterances per hour

Figure 1.b presents the number of regular, redundant and clarificatory device utterances made per navigational decision point. The organization of the data, however, is the same as in Figure 1.a. From the data presented, road routes appear to require less interaction resources than the other types of routes studied. This corroborates the observer statements discussed in sub-section B. in which road routes were said to be slightly easier than the others.

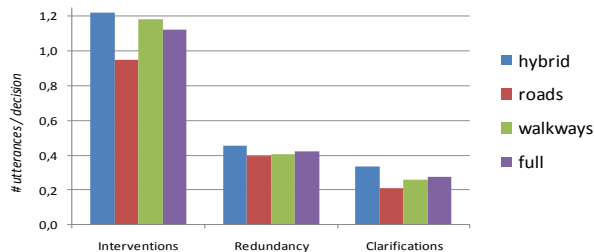


Figure 1.b: System utterances per decision

In the former figure, a large number of redundant utterances can be observed, jeopardizing the likelihood of a favorable analysis of the device with regard to the Maxim of Quantity. The significant number of clarification requests also corroborates this last point with regard to the Maxim of Quantity. Figure 2.a and Figure 2.b below offer this additional analysis by classifying the redundancies produced.

As clearly visible in Figure 2.a, one out of every seven device utterances provided no new information or useful redundancies and, therefore, could have been avoided. Furthermore, almost half of the superfluous utterances were considered annoying by test subjects. As a result, it can be concluded that the navigational device's interactive capabilities do not conform to Grice's Maxim of Quantity.

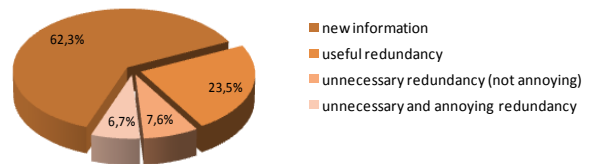


Figure 2.a: Classification of redundancies (% totals)

While of relatively lesser importance, it is nevertheless interesting to note here that Figure 2.b shows that unnecessary and annoying redundancy during walkway routes is greater than unnecessary not annoying redundancy. Therefore, redundancy is more likely to annoy users in walkways. This finding is at odds with the earlier difficulty assessment made by the experiment observer, according to which road routes were said to be easier.

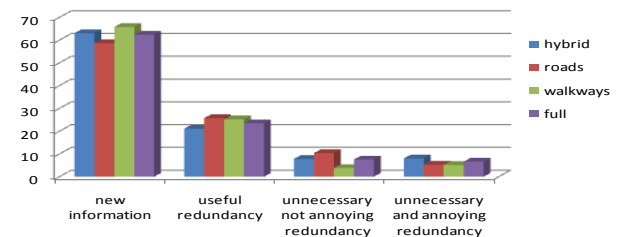


Figure 2.b: Classification of redundancies (% per route)

Regarding the timing of device utterances, Figure 3.a reveals that 16.7% of all utterances were considered ill-timed by users, suggesting a relatively low degree of informational relevancy in these cases. Although the contents of device utterances are, strictly speaking, limited to information relevant to the navigational task, the untimeliness and exaggerated redundancy of utterances may lead one to conclude that the interactive capabilities of the navigation device do not satisfy Grice's Maxim of Relation, either.

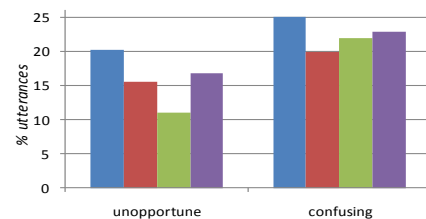


Figure 3.a: Percent of utterances considered ill-timed or confusing

Furthermore, untimely utterances may also endanger the success of the interaction. Figure 3.a shows that users found 23% of all device utterances confusing. With such a high percentage of imprecise or confusing utterances, clearly the Maxim of Manner is not satisfied by the interactive capacities of the navigational device, either. As illustrated in Figure 3.b, untimely and confusing utterances account jointly for the fact that users hesitated in 14% of all decisions made and that 7.7% of all decisions made were incorrect.

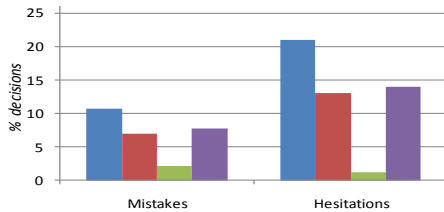


Figure 3.b: Percent of user decisions with mistakes or hesitations

Finally, the graphs demonstrate that walkway routes involve fewer user mistakes and hesitations despite similarly elevated percentages of ill-timed and confusing utterances. Similarly, almost all users appeared to the observer to be more relaxed and self-confident in walkway routes than in other routes. While these results may seem strange given the walkway routes' greater relative difficulty, they can nevertheless be explained by taking into account (1) the higher number of user requests for clarification (see Figures 1.a and 1.b) and (2) the greater insignificance of the environment events (i.e., the low-level task) present in the walkway routes. Consequently and to revise an earlier conclusion, it appears that car traffic in the environment does in fact affect test subjects who utter fewer requests for clarification despite harboring greater doubts regarding the correctness of their navigational decisions. Regarding the hybrid route, users appeared to have problems particularly when moving from a pedestrian walkway to a road. This is understandable when one considers the fact that while continuous, contextual changes affect the user significantly, the non-human-like navigation device is incapable of adapting its interactions to these contextual shifts.

**E. Subjective Results**

The subjective opinions of test subjects regarding their interactions with the navigation devices were gathered in open interviews following the execution of the different route tasks and later analyzed by the researchers. The results obtained are presented below in Figures 4.a to 4.d.

Figure 4.a illustrates that all test subjects thought that the device should have provided more information on at least some occasions, yet very few subjects found a lack of information in device utterances to be the norm. Figure 4.b shows that most users considered that the system produced more utterances than were necessary. While only a third of users characterized the interaction as not redundant, the remaining users believed the interaction to have been unnecessarily redundant, with almost half of whom considering the redundancies to be annoying to the point of jeopardizing the success and tolerability of the interaction.

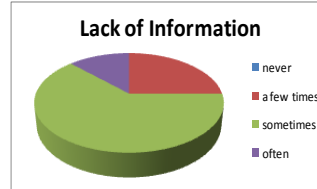


Figure 4.a: How often was information lacking?

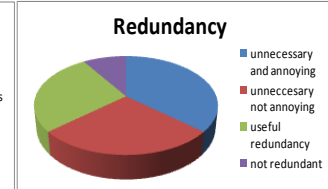


Figure 4.b: Was the interaction redundant?

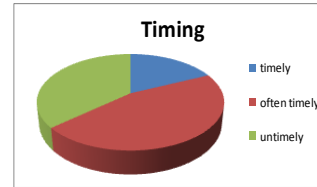


Figure 4.c: Was the information in a timely manner?



Figure 4.d: What caused provided in the mistakes?

Relevancy, on the other hand, can be examined from Figures 4.b and 4.c. User opinions regarding the usefulness of device utterances are particularly low. Not only were interactions viewed as unnecessarily redundant, but most users also believed the devices to have provided ill-timed information (i.e., either too early or too late), with more than a third of whom having believed this untimeliness to be the norm for the devices. It is clear yet again that the interactive capabilities of the devices do not meet the Maxim of Relation.

Regarding manner, 45% of users considered device utterances confusing. However and as shown in Figure 4.d, when asked about the causes of those errors in the navigation process, nearly the same, large percentage of users believed the errors to have been caused either by a lack of clarity or by a lack of information.

Finally, when confronted directly on the question of preference between a human guide or navigation device, the vast majority of test subjects stated a preference for the human guide, with only 9% of subjects stating a preference for the navigation device. Among this 9% of subjects preferring the navigation device, the majority were highly technologically trained individuals who either (1) enjoyed daily experience with similar devices or (2) possessed an engineering degree and spent over 10 hours per day with computers. When asked about the reasons for their expressed preference, over 70% of users mentioned that human beings are more precise and can adapt their utterances to the circumstances of the interaction. Users preferring navigation devices tended to focus on their availability and speed..

**IV. DISCUSSION**

From the objective results presented in the previous section, it is clear that the Gricean Maxims of Quantity and Relation are not met by the interactions of the navigation device. The interaction, however, may still be considered good enough insofar as only one device utterance out of every seven was of no use and less than half of these useless utterances annoyed users. Nevertheless, the interactions lead to frequent user hesitation (in approximately 14% of user decisions made) and multiple user errors (in 8% of user decisions made). This last fact is likely due to the lack of

timely and precise device utterances (one out of every six device utterances was ill-timed and nearly one out of every four device utterances was confusing in some way) revealing the inadequacy of device utterances in meeting the Maxim of Manner. It must be noted that while navigation device interactions may be correctly and satisfactorily developed under optimal conditions, real circumstances are rarely if ever optimal. Thus, it may be concluded that the Gricean Maxims are rarely if ever met by these devices' interactions.

Despite these conclusions, however, defenders of navigation devices might argue that any mistake made may be quickly addressed and corrected by the device through the recalculation of a new route. Furthermore, human interaction is far from perfect, either. While this may be true, however, the impact of the interaction on other concurrently executed tasks is definitive. Additionally, comments recorded during breaks from route executions revealed some novice users to have held unrealistically high expectations for the navigation device and other users to have been prejudiced against technology. Finally, while test subjects knew nothing about the ultimate goal of the experiments – the evaluation of the interactive capabilities of navigation devices – they nevertheless sensed that navigation device technology was being closely examined. As a result and given particular user profiles, some users may have been inclined to more vehemently defend the devices than they would have done otherwise, while other users may have done just the opposite.

#### V. CONCLUSIONS AND AREAS FOR FUTURE RESEARCH

This study has aimed to assess the quality of interactions developed by current navigation devices and to see whether other interactive paradigms could be considered for them. Although some features of human interaction, such as audio channel, have been incorporated into the devices, they are still quite mechanical and primitive compared with fully-human interaction. Among the principal weaknesses of these devices is their inability to adapt to circumstances involving the interlocutor or the interaction itself. This lack of adaptability leads to errors in the quantity of information provided, as well as in the timeliness and manner of the utterances produced. These shortcomings can be overcome through the inclusion of a natural, human-like NI paradigm.

Notice, however, that while this is clearly the case, NI research is not yet mature enough to completely address this necessity. Besides, improvements could still be included in navigation devices as new advances in NI research are made. These gradual, small upgrades should be made while, at the same time, attempting to avoid conflicts resulting from excessively high user expectations. In any case, even if the NI upgrades were complete, the devices would still not likely achieve perfect marks from users. Expert human guides do not generally receive perfect user marks, either.

Concerning the methodology used in the experiments of the present study, it seems adequate for obtaining useful information on the weaknesses of current navigation technologies. Gricean Maxims also proved a good point of reference for the comparison of human and non-human interaction paradigms. However, as interactive capabilities of devices continue to evolve and include more and more

human-like interactive strategies (e.g., flouting the Gricean Maxims), a more sophisticated theoretical background will be necessary for their complete assessment. For this reason, it would be particularly advantageous for future assessments to develop a tailored evaluation methodology for NI systems.

#### ACKNOWLEDGMENT

This study includes results of the research projects SemAnts (TSI-020100-2009-419) and Thuban (TIN2008-02711) funded by the Spanish Ministries of Industry and Education, respectively. Besides, it has been partially supported by the Reg. Government of Madrid under the Research Network MA2VICMR (S2009/TIC-1542).

#### REFERENCES

- [1] Adamczyk, P.D, Bailey, B.P. If Not Now, When? The Effects of Interruption at Different Moments Within Task Execution. *SIGCHI conf. on Human factors*, 271–278, 2004 ACM Press.
- [2] Bailey, B, Iqbal, S. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Trans. on CHI* 14, 4, 2008.
- [3] Brown, P., Levinson, S.C. Politeness: Some Universals in Language Usage. *Cambridge University Press*, 1987.
- [4] Calle, J, García-Serrano, A, Martínez, P. Intentional Processing as a Key for Rational Behaviour through Natural Interaction. *Interacting With Computers* 18, 1419–1446, Elsevier, 2006.
- [5] Calle, J., Martínez, P., Valle, D., Cuadra, D. Towards the Achievement of Natural Interaction. In *Engineering the User Interface: from Research to Practice*, pp. 1–19, Springer 2008
- [6] Dabbish, L. and Kraut, R. Controlling Interruptions: Awareness Displays and Social Motivation for Coordination. In *Proc of CSCW 2004*, pp. 182--191, *ACM Press*.
- [7] Gee, J.P. Introduction to Discourse Analysis. *Routledge* 1999.
- [8] Grice, H.P. (1975). Studies in the Way of Words. *Harvard University Press*, 1989.
- [9] Harris, R.A. Doing Things with Words. *Voice Interaction Design*, pp 75-125. Morgan Kaufmann, 2005.
- [10] Hofs, D, Theune, M, Akke, R. Natural interaction with a virtual guide in a virtual environment: A multimodal dialogue system. *Multimodal User Interf.*, 3 (1-2). pp. 141-153. Springer 2010.
- [11] Horvitz, E. and Apacible, J. Learning and Reasoning About Interruption. *5<sup>th</sup> ACM Int. Conf. on Multimodal Interf.* 2003.
- [12] Iqbal, S.T. and Bailey, B.P. Oasis: A framework for linking notification delivery to the perceptual structure of goal-directed tasks. *ACM Computer-Human Interact.* 17, 4, 2008.
- [13] Janlert, L.E. and Stolterman, E. Complex interaction. *ACM Trans. Comput.-Hum. Interact.* 17, 2, May 2008.
- [14] Kamar, E., Grosz, B.J., and Sarne, D. Modeling User Perception of Interaction Opportunities in Collaborative Human-Computer Settings. *AAAI 2007*, vol. 2, 1872–1873.
- [15] Valle, D., Rivero, J., Calle, J., Cuadra, D. Applying a Methodological Approach to the Development of a NI System. *3<sup>rd</sup> WSKS'10*. Corfú (Greece), 2010.
- [16] Wahlster, W. SmartKom: foundations of multimodal dialogue systems. Berlin, *Springer* 2006

## ***Kalman Filter for Tracking Robotic Arms Using low cost 3D Vision Systems***

*Enrique Martinez Berti*

Instituto de Automática e Informática  
Industrial  
Univesitat Politècnica de València  
Valencia, Spain  
e-mail: enmarbel@etsii.upv.es

*Antonio José Sanchez Salmerón*

Instituto de Automática e Informática  
Industrial  
Univesitat Politècnica de València  
Valencia, Spain  
e-mail: asanchez@isa.upv.es

*Francesc Benimeli*

Instituto de Automática e Informática  
Industrial  
Univesitat Politècnica de València  
Valencia, Spain  
e-mail: frabean@ai2.upv.es

**Abstract**—This paper describes a platform which allows humans to interact with robotic arms using augmented reality. Low cost “kinect” cameras (Xbox 360) are used for tracking human skeletons and locations of robot’s end effectors. The main goal of this paper is to develop robust trackers on this platform. Concretely, a Kalman filter is used for tracking robotic arms using data received from these sensors. It comes to finding a low cost platform for human-robot interactions.

**Keywords**—low cost vision system; Kalman filter; augmented reality; kinematics; Human-machine interaction.

### I. INTRODUCTION

There is a wide range of industrial processes in which robotic systems are present. Nowadays, the required characteristics for such industrial processes are high efficiency, flexibility and adaptability. Human-robotic systems interaction is a key solution to accomplish these requirements, establishing a synergy between the best features of both robots and humans: robot’s precision and high efficiency and human’s flexibility and adaptability.

Human-machine interactions have numerous applications such as assembly tasks [3][12], wheelchair controls through different types of sensors [2][13], developments of servomechanisms [1], developments of intelligent robots [4], and so on.

This paper provides the basis for human-machine interaction in order to increase efficiency in pieces assembly processes, whose flexibility and adaptability characteristics require a close interaction between humans and the robotic systems. Interactions between humans and robots improve the efficiency of complex assembly processes, especially when intelligence is required by the system [3]. However, a precondition for this close relationship is human safety. Many research advances have been carried out in this area and now, some surveillance systems based of sensors used to interact with robots in the market can be found. Intelligent assistance devices (IAD) are the basis for introducing human beings in assembly processes in order to use their cognitive and sensory-motor skills to carry out assemblies with high flexibility.

The main goal of this research consists on creating a platform which can be used as a basis for developing

applications with interactions between humans and machines. A simple practical case of human-robot interaction has been implemented to check this platform.

#### A. Related Research

We consider the problem of estimating and tracking 3D configurations of complex articulated objects from images, e.g., for applications requiring 3D robot arms pose, human body pose and hand gesture analysis. There are two main schools of thought on this. Model-based approaches presuppose an explicitly known parametric articulated object model and estimate the pose either by directly inverting the kinematics (which has many possible solutions and which requires known image positions for each part [26]) or by numerically optimizing some form of model-image correspondence metric over the pose variables, using a forward rendering model to predict the images (which is expensive and requires a good initialization, and the problem always has many local minima [24]). An important subcase is model-based tracking, which focuses on tracking the pose estimate from one time step to the next starting from a known initialization based on an approximate dynamical model [17][23]. In contrast, learning-based approaches try to avoid the need for explicit initialization and accurate 3D modeling and rendering, instead capitalizing on the fact that the set of typical articulated object poses is far smaller than the set of kinematically possible ones and learning a model that directly recovers pose estimates from observable image quantities. In particular, example-based methods explicitly store a set of training examples whose 3D poses are known, estimating pose by searching for training image(s) similar to the given input image and interpolating from their poses [15][19][22][25].

There is a good deal of prior work on articulated objects pose analysis, but relatively little on directly learning 3D pose from image measurements. Brand [16] models a dynamical manifold of human body configurations with a Hidden Markov Model and learns using entropy minimization, Athitsos and Sclaroff [14] learn a perceptron mapping between the appearance and parameter spaces, and Shakhnarovich et al. [22] use an interpolated-k-nearest-



neighbor learning method. Human pose is hard to ground truth, so most papers in this area [14][16][19] use only heuristic visual inspection to judge their results. However, Shakhnarovich et al. [22] used a human model rendering package (POSER from Curious Labs) to synthesize ground-truthed training and test images of 13 degrees of freedom upper body poses with a limited ( $\pm 40^\circ$ ) set of random torso movements and viewpoints. Several publications have used the image locations of the center of each body joint as an intermediate representation, first estimating these joint centers in the image, then recovering 3D pose from them. Howe et al. [18] develop a Bayesian learning framework to recover 3D pose from known centers, based on a training set of pose-center pairs obtained from resynthesized motion capture data. Mori and Malik [19] estimate the centers using shape context image matching against a set of training images with pre-labeled centers, then reconstruct 3D pose using the algorithm of [26]. These approaches show that using 2D joint centers as an intermediate representation can be an effective strategy.

With regard to tracking, some approaches have learned dynamical models for specific human motions [20][21]. Particle filters and MCMC methods have been widely used in probabilistic tracking frameworks, e.g., [23][27]. Most of these methods use an explicit generative model to compute observation likelihoods.

### B. Overview of the Approach

Using the same philosophy as for tracking human skeleton, former approaches can be applied to track robot arms. We propose to use a low cost vision system which requires a discrete Kalman filter. This allows tracking joint variables at each instant of time.

### C. Organization

Section 2 describes the global system. Section 3 describes the low cost 3D vision system. Section 4 describes the Kalman filter used to track a robot arm. Section 5 presents the robot used. Section 6 describes a human skeleton tracking approach. Finally, Section 7 concludes with some discussions and directions of future work.

## II. GLOBAL SYSTEM DESCRIPTION

A platform for human-machine interaction using augmented reality [8] has been performed between robotic systems and human beings. This platform is a distributed system where processes can communicate easily between them. The main functionalities offered by this platform are:

- 1) Communications between processes via XML [5].
- 2) Safety controls.
- 3) Tracking of robot arm poses.
- 4) Tracking of human skeletons.
- 5) Handle of augmented reality scenes.

A practical case of human-robot interaction has been implemented to check this platform. Figure 1 shows the

distribution of the physical components (robot and camera) of this application.



Figure 1. Low Cost 3D Vision System.

In this case, the distributed system is composed by two processes to perform interactions between a human and a robot. One process realizes the monitoring of human skeleton. The other process realizes the monitoring of the robot arm pose. The data information for human and robot state estimation is obtained by a low-cost 3D camera.

The XML messages set developed ad-hoc for an application use special communication software. This software is called RT-SCORE [6] and it is a system (based on a blackboard system) that allows to assign a communication channel between processes. The “channel” concept is similar to a “hall” in a chat communication system (the chat communication is RT-SCORE); so, only the entities connected to a channel receive the information sent into this channel.

Safety regulations require introducing guardrails, so that humans do not have direct access to industrial robots workspaces. To achieve a human-robot interaction a safety protocol has been established that allows such interaction without risk of serious damages. The safety control system calculates the human and robotic arms location, so that the closer they get, the slower the robot moves.

MatLab [10] has been used in order to implement both processes.

The robot arm monitor tracks the end effector and joint angles of the robot. However, it is needed a Kalman filter for tracking robot arm poses.

The human skeleton monitor uses third party libraries with functions to estimate locations of each body part. Some human skeleton poses are used to handle virtual objects in the augmented reality scenario. Additionally these virtual objects can be shown on real RGB images captured by the camera.

## III. LOW COST 3D VISION SYSTEM

The vision system used in this practical case is the “Kinect” camera, which consists of two optical sensors whose interaction allows a three-dimensional scene analysis. One of the sensors is an RGB camera which has a video resolution of 30 fps. The image resolution given by this camera is 640x480 pixels. The second sensor has the aim of obtaining depth information corresponding to the objects

found at the scene. The working principle of this sensor is based on the emission of an infrared signal which is reflected by the objects and captured by a monochrome CMOS sensor. A matrix is then obtained which provides a depth image of the objects in the scene, called DEPTH.

The calibration process of this camera can be seen in [9]. Calibration is needed to relate both camera and robot coordinates reference systems. Therefore objects located by the camera can be handled by the robot.

#### IV. KALMAN FILTER

The Kalman filter [11] is used in sensor fusion and data fusion. Typically real time systems produce multiple sequential measurements rather than making a single measurement to obtain the state of the system. These multiple measurements are then combined mathematically to generate the system's state at that time instant.

Data fusion using a Kalman filter can assist computers to track objects in videos with low latency (not to be confused with a low number of latent variables). The tracking of objects is a dynamic problem, using data from sensor and camera images that always suffer from noise. This can sometimes be reduced by using higher quality cameras and sensors but can never be eliminated, so it is often desirable to use a noise reduction method.

The iterative predictor-corrector nature of the Kalman filter can be helpful, because at each time instance only one constraint on the state variable needs to be considered. This process is repeated considering a different constraint at every time instance. All the measured data are accumulated over time and help in predicting the state.

Video can also be pre-processed, using a segmentation technique, to reduce computation and hence latency.

The discrete Kalman filter [11] is implemented as follows:

- 1) State prediction:

$$\hat{X}_t^* = A\hat{X}_{t-1} \tag{1}$$

- 2) Prediction of error covariance:

$$P_t^* = AP_{t-1}A^T + Q \tag{2}$$

- 3) Calculate the constant gain  $K$ :

$$K_t = P_t^*H^T(HP_t^*H^T + R)^{-1} \tag{3}$$

- 4) Update:

$$\hat{X}_t = \hat{X}_t^* + K_t(Z_t - H\hat{X}_t^*) \tag{4}$$

- 5) Update error covariance:

$$P_t = (I - K_tH)P_t^* \tag{5}$$

The Kalman filter has been applied to depth information. The values returned by depth images are not always right. This happens because the sensor does not detect the depth correctly when the infrared light is not properly reflected on the object. In this case, the input value to the Kalman filter is the depth value of  $z_d$  (state) corresponding to the distance between the camera and the object. In (1) and (2) the  $z_d$  value and covariance is predicted to the next step. Equations

(3), (4) and (5) are the equations to correct the discrete Kalman filter. In (3), a new gain of Kalman is calculated. Equations (4) and (5) calculate a new value of  $z_d$  predicted, and new covariance of error, respectively.

Three Kalman filters have been implemented: one for each of the three points used to locate (position and orientation) the end effector. Figure 2 shows a pose estimated during the robot movement. It can be seen the three points detected on the end effector.

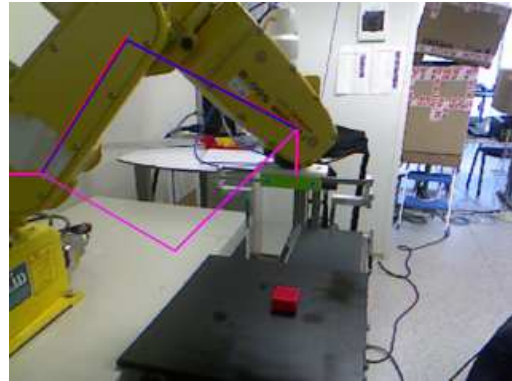


Figure 2. Robot arm pose tracking.

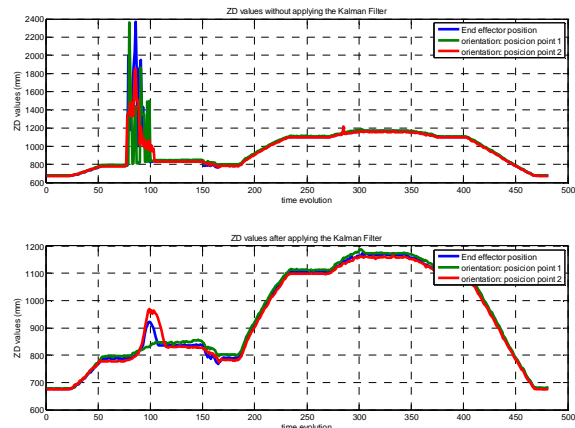


Figure 3. Kalman Filter evolution.

Figure 3 shows the evolution of depth information for comparing results obtained with Kalman filter and without applying the Kalman filter. The first graph shows parameter  $z_d$  over time without applying the Kalman filter. The second graph shows  $z_d$  over time applying the Kalman filter. At time 100, incorrect  $z_d$  values can be observed when not applying the filter because the camera does not get properly information. It can be seen that the Kalman filter makes a correction of these values.

#### V. ROBOT FANUC 200IB CONTROL

DH (Denavit-Hartenberg) is used to solve direct and inverse kinematics problems. Figure 4 shows coordinate systems and articulation axes used for this Fanuc robot.

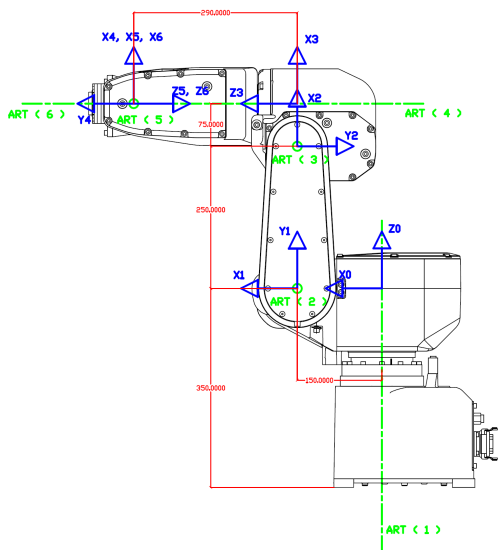


Figure 4. Coordinate systems and articulation axes.

Hence, the following DH parameter table is obtained:

TABLE 1. DENAVIT-HARTEMBERG.

ART	$\theta$	$d_i$	$a_i$	$\alpha_i$
1	0°	0	150	90°
2	90°	0	250	0
3	0°	0	75	90°
4	0°	290	0	90°
5	0°	0	0	90°
6	0°	0	0	0

The direct and inverse kinematics problems are solved using these parameters. These kinematic models allow tracking the robot by using the kinect, so that the end effector position is identified on the image and the state of the robot joints is calculated. The Kalman filter is necessary to filter the information captured by vision sensor.

VI. MONITORING HUMAN SKELETON

To monitor the skeleton of a human being, the toolbox [7] and [8] has been used in Matlab. The NITE tracking human body module aim is based on extracting the most important features of the human skeleton and following them over time. Figure 5 shows the result of the skeleton detection by the kinect.

Tracking the human skeleton is necessary to control the actions in order to interact with the robot. In the present case study, it is used to insert virtual parts in the scene. A virtual piece will be created on the hand using a set gesture and with another set gesture the object will be fixed in that position. Figure 6 shows the gesture to pick up a virtual object and Figure 7 shows the gesture to place the object in a fixed position.

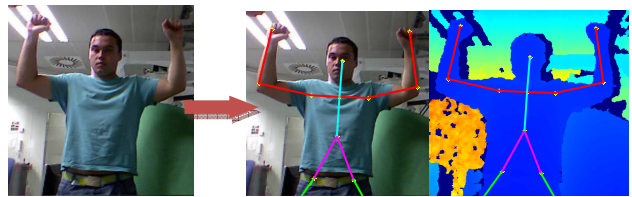


Figure 5. Skeleton pose control.

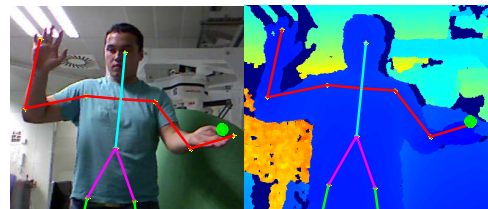


Figure 6. Take object.

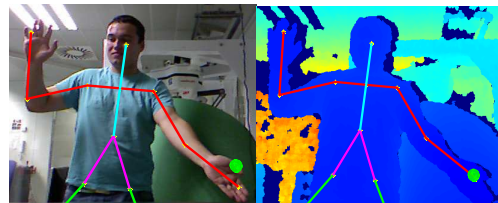


Figure 7. Place object.

The act of creating a virtual object is performed by placing the arm and forearm in an angle of 90 degrees. Once the piece appears on the image, it follows the arm movements until the subject performs the gesture for placing the piece, which consists on stretching out the arm to the desired position.

VII. CONCLUSION

A platform that serves as the basis for developing applications which establish interaction between humans and robots has been created.

Using this platform we have carried out a simple case study of interaction between a human being and a robotic system that allows the handle of virtual and real parts between a human and a robot.

A discrete Kalman filter is used to reduce noise in data get it from the low cost vision system which allows tracking robot arms. A human body can be modeled like some interconnected robots. Therefore this method can be extrapolated for tracking human skeletons.

In a future work, we will explore new methods to track articulated objects based on efficient robot models, like screw theory instead of DH model.

ACKNOWLEDGMENT

This work was partially funded by the Universitat Politècnica de València research funds (PAID 2010, Project no. 2566) and by Spanish government and the European Community under the project DPI2010-20814-C02-02 (FEDER-CICYT) and DPI2010-20286 (CICYT).

REFERENCES

- [1] S. Domínguez, E. Zalama and J. Gomez (2005) "Desarrollo de una cabeza robótica con capacidad de seguimiento visual e identificación de personas" XXVI Jornadas de automática.
- [2] J. Gonzalez, C. Galindo, J.A. Fernandez, J.L. Blanco, A. Muñoz and V. Arevalo (2008) "La silla robótica SENA. Un enfoque basado en la interacción hombre-máquina", Revista Iberoamericana de Automática e Informática Industrial, pp. 38-47.
- [3] J. Kruger, T.K. Lien and A. Verl (2009) "Cooperation of human and machines in assembly lines" CIRP Annals-Manufacturing Technology, pp. 628-646.
- [4] F. Ramirez (2010) "Avances en interacción hombre-máquina" ECIPERU Revista del encuentro científico internacional pp. 29-36.
- [5] E. Olmos, M. Bosch and A. Sánchez (2009) "Programación de Robots a través de XML", XXX Jornadas de Automática, Valladolid, España.
- [6] J.L. Posadas, P. Pérez, J.E. Simó, G. Benet and F. Blanes (2002). "Communications Structure for Sensory Data in Mobile Robots". Engineering Applications of Artificial Intelligence, vol. 15, no. 3, pp. 341-350.
- [7] Prime Sensor™ NITE 1.3 Algorithms notes. <http://www.primesense.com>.
- [8] PrimeSense™ NITE Controls User Guide. <http://www.primesense.com>.
- [9] E. Martinez Berti, D. Hernandez Campos, and A. Sánchez Salmerón, "Visión 3D de Bajo Coste para la interacción humano-robot utilizando filtro de partículas". Jornadas de Automática, Sevilla 2011.
- [10] MatLab: <http://www.mathworks.es/products/matlab/index.html> (November, 2011)
- [11] Greg Welch and Gary Bishop, "An Introduction to the Kalman Filter", Design 7, n.º. 1 (2001): 1-16
- [12] T. Itoh, K. Kosuge, and T. Fukuda, "Human-machine cooperative telemanipulation with motion and force scaling using task-oriented virtual tool dynamics", IEEE Transactions on Robotics and Automation 16, n.º. 5 (October 2000): 505-516
- [13] F. Leishman, O. Horn, and G. Bourhis, "Smart wheelchair control through a deictic approach", Robotics and Autonomous Systems 58, no. 10 (October 31, 2010): 1149-1158
- [14] V. Athitsos, and S. Sclaroff, "Inferring Body Pose without Tracking Body Parts," Proc. Int'l Conf. Computer Vision and Pattern Recognition, 2000.
- [15] V. Athitsos, and S. Sclaroff, "Estimating 3D Hand Pose from a Cluttered Image." Proc. Int'l Conf. Computer Vision, 2003.
- [16] M. Brand, "Shadow Puppetry," Proc. Int'l Conf. Computer Vision, pp. 1237-1244, 1999.
- [17] C. Bregler, and J. Malik, "Tracking People with Twists and Exponential Maps," Proc. Int'l Conf. Computer Vision and Pattern Recognition, pp. 8-15, 1998.
- [18] N. Howe, M. Leventon, and W. Freeman, "Bayesian Reconstruction of 3D Human Motion from Single-Camera Video," Neural Information Processing Systems, 1999.
- [19] M. Mori, and J. Malik, "Estimating Human Body Configurations Using Shape Context Matching," Proc. European Conf. Computer Vision, vol. 3, pp. 666-680, 2002.
- [20] D. Ormoneit, H. Sidenbladh, M. Black, and T. Hastie, "Learning and Tracking Cyclic Human Motion," Neural Information Processing Systems, pp. 894-900, 2000.
- [21] V. Pavlovic, J. Rehg, and J. MacCormick, "Learning Switching Linear Models of Human Motion," Neural Information Processing Systems, pp. 981-987, 2000.
- [22] G. Shakhnarovich, P. Viola, and T. Darrell, "Fast Pose Estimation with Parameter Sensitive Hashing," Proc. Int'l Conf. Computer Vision, 2003.
- [23] H. Sidenbladh, M. Black, and L. Sigal, "Implicit Probabilistic Models of Human Motion for Synthesis and Tracking," Proc. European Conf. Computer Vision, vol. 1, 2002.
- [24] C. Sminchisescu, and B. Triggs, "Kinematic Jump Processes for Monocular 3D Human Tracking," Proc. Int'l Conf. Computer Vision and Pattern Recognition, June 2003.
- [25] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla, "Filtering Using a Tree-Based Estimator," Proc. Int'l Conf. Computer Vision, 2003.
- [26] C. Taylor, "Reconstruction of Articulated Objects from Point Correspondences in a Single Uncalibrated Image," Proc. Int'l Conf. Computer Vision and Pattern Recognition, 2000.
- [27] K. Toyama, and A. Blake, "Probabilistic Tracking in a Metric Space," Proc. Int'l Conf. Computer Vision, pp. 50-59, 2001.

# Predicting Performance and Situation Awareness of Robot Operators in Complex Situations by Unit Task Tests

Tina Mioch, Nanja J. J. M. Smets, Mark A. Neerincx  
*TNO*  
 Kampweg 5, 3769 DE Soesterberg, The Netherlands  
 {tina.mioch, nanja.smets, mark.neerincx}@tno.nl

**Abstract**—Human-in-the-loop field tests of human-robot operations in high-demand situations provide serious constraints with respect to costs and control. A set of relatively simple unit tasks might be used to do part of the testing and to establish a benchmark for human-robot performance and situation awareness. For an urban search and rescue (*tunnel accident*) scenario, we selected and refined the corresponding unit tasks from a first version of a test battery. First responders (fire-men) conducted these unit tasks with a state-of-the-art robot and, subsequently, had to perform the *tunnel accident* mission in a realistic field setting with the same robot. The *Detect objects* unit task proved to partially predict operator’s performance and the operator’s collision awareness in the scenario. Individual differences, particularly age, had a major effect on performance and collision awareness in both the unit tasks and scenario.

**Keywords**—Human-robot cooperation; Performance evaluation

## I. INTRODUCTION

Unmanned Ground Vehicles (UGVs) are intended to be deployed in diverse, high-demanding environments. Human-robot team performance is often critical (e.g., time pressure, high error costs), and dependent on team’s skills to cope with the dynamic situational conditions, for example in the urban search and rescue domain. Evaluation of the robots before actual deployment is of utmost importance, but the opportunities to conduct realistic field experiments are constrained due to the limited availability of end-users and test sites. Furthermore, objective evaluation poses fundamental difficulties due to the ‘situatedness’ of robots’ effectiveness and efficiency so that outcomes may be hard to generalize.

In this study, we investigate the applicability and validity of a usage-centered evaluation methodology for unmanned ground vehicles. This evaluation methodology provides standard task assignments and metrics on human-robot collaboration. The idea is that a set of relatively simple and abstract unit tasks can be used to assess basic aspects of this collaboration and to establish a benchmark for human-robot performance and situation awareness. Such tests can decrease the need for evaluating the human-robot performance in the environment in which it will actually be deployed. The assumption is that these tests predict the performance

in a realistic scenario for an important part. The application of the proposed test battery with ‘unit tasks’ should help

- to generalize,
- to standardize (compare results of different tests), and
- to interpret outcomes in terms of the robot’s functional components.

For a more detailed motivation and positioning in a usage-centered UGV evaluation and design methodology, see [1]. It should be noted that the emphasis of this research lies on a first evaluation of the applicability and validity of the methodology. Our approach is to instantiate the methodology for one particular research question, namely human-robot collaboration in an urban search and rescue scenario (“tunnel accident”). For each task in the test battery, the interaction of the whole system, meaning one robot together with its operator, is evaluated. The test battery tasks are not intended to do isolated tests of specific robot technologies or performance tests of either the individual robot or individual operator (i.e., the focus is on joint human-robot operation).

The following research questions can be identified:

- Is the performance and situation awareness of the participants in the test battery a good prediction of the performance and situation awareness in the scenario?
- Can the unit tasks help to explain operator performance in complex scenarios?

Individual differences can have a major effect on operational outcomes. To get first insight in such effects, we will analyze whether individual factors such as spatial ability and experience in computer games influence the performance and situation awareness of the operator, and whether these effects are similar for the test battery and scenario setting.

The paper is structured as follows: first, we will describe how this research can be placed in the context of performance evaluation for human-robot cooperation, followed by a description of the method to answer the research question. Subsequently, the results of the experiment are given and discussed.

## II. BACKGROUND

In this section, it is described how this research can be placed in the context of performance evaluation for human-robot cooperation.

### A. Situated Cognitive Engineering Methodology

To establish the set of functional requirements with the corresponding metrics for evaluation, the situated Cognitive Engineering (sCE) Methodology [2] is applied. Following the sCE methodology, the operational demands, human factors knowledge and technological constraints were analyzed and used to specify design scenarios and a requirements baseline. An example of a requirement is given in Figure 1. The requirements baseline consists of claims that justify the requirements, and use cases that contextualize and organize these requirements.

Requirement 1.1	The robot should be able to be manually controllable by an operator	
Claim 1	If the robot needs navigational help (e.g. getting stuck or precise maneuvering), the operator can take manual control over the robot	
	+	The efficiency of the robot and operator cooperation will increase if the operator is capable to take over control for situations in which the operator thinks that manual control is needed
	-	Manually controlling the robot will take up attention of the operator, increasing the experience of the workload
		Measuring the time it takes for the operator and the robot to finish a task in which the robot gets stuck
		Question(s) about the amount of workload the operator is/was experiencing when teleoperating
Use cases	UC_0004(Main success scenario: step 1, 2, 5)	
Task battery tasks	T1: Stop before collision	Tests manual control of operator over the robot and a good awareness about speed and distance to the object

Figure 1. An example of a requirement, with a claim and the corresponding unit task.

Subsequently, we identified unit tasks in the test battery set, which addressed these requirements. This means that to execute the unit task successfully, the requirement must be met, just as this should be the case for the scenarios implementing the use cases. For each requirement, at least one unit task that manifests this requirement in the scenario was selected.

### B. Performance evaluation

Several categories of human-robot cooperation metrics can be distinguished: general metrics, collaboration, and user interfaces. In this paper, we concentrate on the general performance metrics. These include for example efficiency, effectiveness, task load and emotions, and situation awareness. In the following, the predictability for general performance and situation awareness is analyzed.

In general, experiment setups for evaluations can differ in the dimensions *fidelity* and *realism* [3]. *Fidelity* expresses how close the collaborative operations resemble the actual "rules" of operations and their internal and external dependencies (i.e., the social and environmental dependencies). *Realism* specifies whether the evaluation environment is represented realistically ("Does it look, feel and smell like

a disaster?"), for example from low realism in a virtual environment, to a high realism in an earthquake site.

Different experimentation environments have different advantages and disadvantages. Evaluating a robot in a "real disaster site" for example has high realism and high fidelity, but it is costly. Furthermore, there is lack of controllability and you cannot test all kind of settings without the risk for damage or injuries. Therefore, specific test arenas are being set up, such as NIST, which have different levels of realism [4]. However, fidelity may remain somewhat lower, because the rescue team cannot operate conform their complete set of coordination and collaboration policies.

As a complementary approach, we propose to identify unit tasks that resemble basic functionality of human-robot collaboration in envisioned scenarios. The higher the resemblance, the higher the fidelity. Here, we will focus on the collaboration between two actors, the robot and operator, however, this approach can be extended to more actors. Subsequently, these tasks are applied to test the collaboration in a controlled setting (preferably with the same environmental constraints as the real setting). In this paper, we evaluate whether the human-robot performance in a test battery can predict actual performance in a field test. For a more extensive motivation, overview and placement of the test battery in comparison to other evaluation environments, see [1]. The field test performed in this study has a high realism.

## III. METHOD

In this section, the method is described in detail.

### A. Task

As described in Section II-A, the unit tasks were selected by requirements matching. The experiment consisted of two parts, namely the test with a selection of tasks from the test battery, and the test with the scenario. The following unit tasks were selected:

- *Detect objects in the environment.* The robot is placed at the entrance of a room. In the room, several warning signs printed on A4 paper can be found. The participant has to find the signs, and situate them on a map, with a time limit of two minutes.
- *Slalom.* The participants have to drive slalom around pylons as fast as possible without touching the pylons.
- *Move through narrow hallway.* The participants have to drive through a narrow hallway as fast as possible without touching the walls.
- *Stop before collision.* At the end of the hallway, participants have to maneuver the robot as close as possible to the wall, without touching the wall.

The second part of the experiment was the execution of the scenario. The scenario was a car accident in a tunnel. The situation in the tunnel was not clear, and more information was needed. There was smoke development in the tunnel.

A robot, controlled by the participants, was deployed to gather information. The participants were asked to answer the following questions:

- Are there cars in the tunnel? If so, where are these?
- How is the layout of the situation?
- Are there victims? And if there are, how many were there, and where?
- Look for fire and dangerous substances, depicted by pictures of warning signs.

While navigating through the scenario area, participants had to indicate on a whiteboard what they saw, by using magnetic icons and whiteboard marker. The magnetic icons were: pallet, truck, warning sign for fire, warning sign for dangerous substance, car, barrel, victim and a cardboard box.

### B. Design

The experiment was within subject, and each participant first performed the test battery tasks, followed by the scenario.

### C. Materials

The following materials were used in the experiment:

- An unmanned ground vehicle, the *Generaal* (see Figure 2), has been custom-made at TNO in Soesterberg and has been used in other studies as well. For a detailed description, see [5]. The vehicle has been specifically designed for telepresence control, with a pan-tilt-roll unit with a camera system mounted on top of it. The telepresence control station consists of a head-tracking head-mounted display (HMD) (see Figure 3), a steering wheel and an accelerator. The head-tracker directs the pan-tilt-roll unit, and the HMD displays the sensor images. This gives the operator the experience of naturally looking around at the remote location. Vehicle control is facilitated by two ‘antennas’ at the side of the robot. These indicate the width of the vehicle as well as the front of the vehicle.
- Hall with separate area for test battery tasks and scenario.
- For setting up the scenario we used the following items: three cars, one motor, five dummy victims, three barrels and three ‘danger’ signs.

1) *Participants*: Nine male participants took part in the experiment as volunteers. All participants were firemen from the fire department of the city Dortmund with an average age of 34. The mean number of years the participants had a driver’s license was 18.

### D. Measures

The following measures were taken during the execution of the test battery tasks and the scenario:

- 1) Performance data
  - Time to finish task

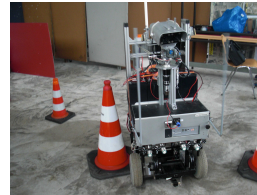


Figure 2. *Generaal* robot of TNO



Figure 3. Head-mounted display interface of *Generaal* robot

- Number of collisions
- 2) Situation Awareness
    - Number of correctly identified objects
  - 3) Performance Perception
    - Perceived collisions
  - 4) Personal characteristics

In the following, we will analyze the performance data, the situation awareness, and the operator’s perception on the performance to determine, whether for these metrics, the test battery is a predictor for the field test measures. The information gained about the personal characteristics is also analyzed.

### E. Procedure

At the beginning of the experiment participants were given a general, written instruction about the experiment. Then a spatial ability test was conducted. Then participants had to fill in a general questionnaire about their background, computer and game experience. An extensive training was conducted with afterwards a learnability questionnaire. Then the participant performed the test battery tasks, with after each test battery task a questionnaire. Then the scenario was performed with a workload questionnaire and map drawing during the scenario, followed by several scenario related questionnaires. The experiment ended with an end questionnaire.

## IV. RESULTS

As depicted in Figure 4, we performed several analyses. First, we performed correlation analysis and multiple regression analysis for performance, situation awareness, and performance perception measures for the scenario, with the performance, situation awareness, and performance perception of the unit tasks as predictor variables (arrow A in Figure 4). In addition, multiple linear regression analyses were performed for the unit tasks and the scenario based on the following predictor variables: age, the amount of kilometers the participant drives per year, and the experience with computer gaming (see arrow B and C in Figure 4). We

decided to use age as a predictor variable and not the number of years the participants had their driver’s license, because some participants did not fill in the question correctly.

**Performance:** For both the unit tests and the scenario, as performance measure, we analyzed the number of collisions. The time it took to finish a task was measured for some of the unit tasks, but not for the scenario, as the operators were given 15 minutes to finish the scenario.

**Situation awareness:** As mentioned above, the operator drew a map of the environment of the scenarios and the test battery tasks. As situation awareness measure, the number of correctly identified objects was analyzed.

**Performance perception:** To measure the performance perception, we selected the measure of collision awareness, as this measure was most practical in defining and applicable for all test battery tasks. For both the unit tests and the scenario, the awareness of the operator of having collided with an object was measured as the difference between the actual number of collisions and the number of collisions reported by the participant.

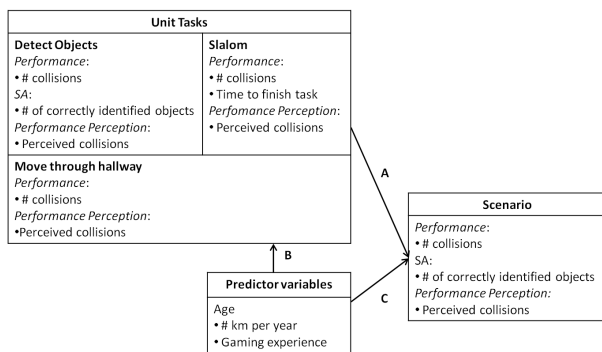


Figure 4. Overview of the analysis

**A. Analysis of the predictive power of the unit task performance for the scenario performance**

One of the questions we want to answer is in how far the unit task performance can predict the performance in the scenario, see arrow A in Figure 4. We are analyzing this for the performance measures (the number of collisions), the SA measure (the number of correctly identified objects), and the operator’s collision awareness.

**Performance:** We conducted a correlation analysis on the performance measure. There was a positive correlation (trend) for the number of collisions, i.e., when a participant collided more in the test battery task *Detect Objects*, the participant also collided more in the scenario, with  $r = 0.44$ ,  $p = 0.063$  (for the scatterplot, see Figure 5).

**Situation awareness:** The correlation for the number of found objects in the *Detect Objects* test battery task and the scenario was not significant. Of the task battery tests, the number of objects found in the *Detect objects* task explains 24 % of the variance in the scenario (see Table I).

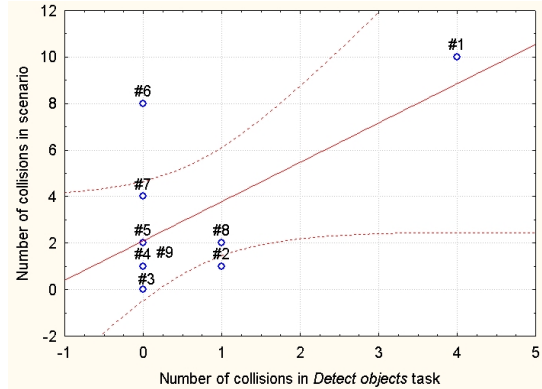


Figure 5. Scatter plot of the performance measure, number of collisions per participant in the *Detect objects* task and the scenario.

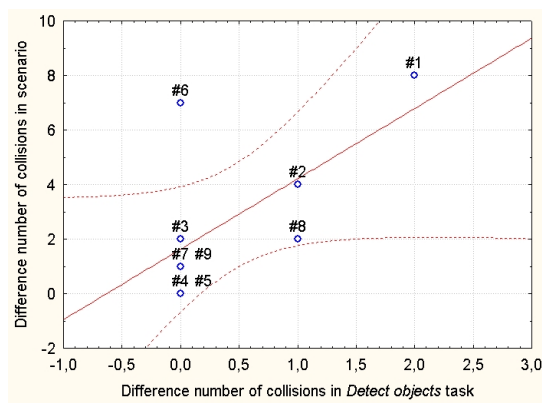


Figure 6. Scatter plot of the performance perception measure, difference between the actual number of collisions and the number of collisions reported per participant in the *Detect objects* task and the scenario.

**Performance perception:** Correlation analysis showed a positive trend between the operator’s collision awareness in the unit task *Detect objects* and the collision awareness in the scenario,  $r = 0.64$ ,  $p = 0.066$ . When there was a larger difference in the actual number of collisions and the number of collisions reported in the *Detect objects* task, this was also the case in the scenario, see Figure 6. When performing a multiple linear regression analysis with the task battery tests, the difference in the number of collisions in the *Detect objects* task explains 40 % of the variance in the scenario (see Table I).

Table I  
PERCENTAGE OF EXPLAINED VARIANCE THE UNIT TASK *Detect objects* ADDS FOR THE SCENARIO.

Criterion	Explained variance R2 (%) by the three predictor variables
Number of objects found in scenario	Number of objects found <i>Detect objects</i> = 24%
Difference in number of collisions scenario	Difference in number of collisions <i>Detect objects</i> = 40%



Table II

PERCENTAGE OF EXPLAINED VARIANCE FOR THE PERFORMANCE AND SA MEASURES THAT THE DIFFERENT PREDICTOR VARIABLES ADD FOR THE UNIT TASKS AND THE SCENARIO.

Criterion	Explained variance R2 (%) by the three predictor variables
Number of objects found in <i>Detect Objects</i>	Age = 32%
Number of objects found in <i>scenario</i>	Age = 74% Add kilometers per year = 86% Add Gaming experience = 89%
Number of collisions in <i>Detect Objects</i>	Age = 13%
Number of collisions in <i>Narrow Hallway</i>	Kilometers per year = 57% Add age = 72%
Number of collisions in <i>Slalom</i>	Age = 59%
Number of collisions in <i>scenario</i>	Age = 38%

Table III

PERCENTAGE OF EXPLAINED VARIANCE FOR THE OPERATOR'S COLLISION AWARENESS THAT THE DIFFERENT PREDICTOR VARIABLES ADD FOR THE UNIT TASKS AND THE SCENARIO.

Criterion	Explained variance R2 (%) by the three predictor variables
Difference in number of collisions, <i>Detect Objects</i>	Age = 20%
Difference in number of collisions, <i>Slalom</i>	Kilometers per year = 45% Add gaming experience of collisions, = 60%
Difference in number <i>Move through narrow hallway</i>	Kilometers per year = 39% Add age = 52% ;
Difference in number of collisions <i>scenario</i>	Age = 37% Add kilometers per year = 60%

*B. Effect of individual differences on the unit task performance and scenario*

In this section, it is analyzed in how far individual differences effect the performance in the unit tasks and in the scenario (see arrow B and arrow C in Figure 4, respectively.) A multiple linear regression analysis was performed to predict the different measures based on the following predictor variables: age, the amount of kilometers driven per year, and the experience with computer gaming.

*Performance and Situation awareness:* Table II shows that age explains most of the variance for the test battery and the scenario. In the regression, it explains the largest part of the variance percentage-wise for all performance variables, of which two are significant (for the number of objects found in the scenario and number of collisions in the slalom). In the scenario, the number of kilometers driven per year and gaming experience is also of influence for the number of objects found. The number of collisions in the *narrow hallway* task is influenced by the amount of kilometers driven per year by the participant.

*Performance perception:* Table III shows that the age of the participants explains the variance percentage-wise for three out of four variables, in the scenario it is significant. Kilometers driven per year also explains the variance for three out of four variables, and is significant in the slalom task. In the slalom task, game experience is of influence as well.

V. DISCUSSION AND CONCLUSION

This study tested a recent method for the evaluation of human-robot collaboration with unit tasks [1]. The *Detect objects* unit task proved to partially predict operator's performance and the operator's collision awareness in the scenario. Individual differences, particularly age, had a major effect on performance and collision awareness in both the unit tasks and scenario.

It should be noted that the *Detect objects* task was the most comprehensive task; both the operational demand of transiting with the robot and observing the environment are included, whereas the other unit tasks are mostly transiting tasks. Hence, the *Detect object* task is the closest of all tasks to the scenario task, in which also transiting and observing the environment. Conversely, if the scenario would have had as main operational demand transiting around the environment, the other unit tasks possibly would have predicted the scenario outcomes better. Our study suggest that, when applying the methodology, the tasks that are used for predicting the performance in the scenario should address the concurrent operational demands.

In addition to the deficient mapping of operational demands on the two "other" unit tasks, effects may have been hidden due to some deficiencies in the amount and property of the data. As in most field studies with real end-users, the number of participants available was limited. In addition, the performance measures of the unit tasks proved not to match perfectly with the scenario measures. For example, the slalom task had two performance measures: the time it took to finish and the number of collisions with the cones. In the scenario, only the number of collisions was relevant, and the time, even though it was limited, was given as a constraint and not as a performance measure. Consequently, the measure *number of collisions* was different in the slalom task compared to the scenario, as the time the task execution took probably influenced the number of collisions. In general, the evaluation measures in the scenario proved to be quite difficult to establish and to incorporate in the unit task measures. Based on the experiences in this test, we will refine the measures in the next tests.

We can further conclude that the unit tasks can be used to explain some operators' performances. As they are specified with a particular challenge in mind, e.g., operational control of the robot, or gaining situation awareness, the reason for a bad or good performance is more easily inferred than when evaluating the scenario performance. For example, because of the *Stop before collision* task, we could determine that the perception of distance was not very good, and that this was the main reason for the number of collisions, instead of difficulty of maneuvering. In general, individual differences, particularly age, proved to have a major effect on performance and situation awareness in both the unit tasks and scenario. Unit tasks show the effects of these

differences and can be of help to see whether higher levels of robot autonomy and advanced situation awareness support can help to decrease problems of some users with current robot control and perception.

#### A. Observations

An interesting observation concerns the performance of participant 6, who consistently showed a deviation from the performance patterns of the other participants. He performed average on the test battery tasks, but clearly below average in the scenario. His perception of his own performance proved to deviate from his actual performance: he most often did not notice the collisions. Probably, he became somewhat overreliant, overestimated his own capabilities, and, consequently, performed worse in the scenario. Without participant 6, the main results of this experiment showed the same pattern, but the level of significance of the effects proved to increase (i.e., the correlations were significant at  $p < 0.5$  without participant 6).

When executing the scenario, several participants believed, after about 12 minutes, that they had explored the whole environment well. After being told that they could go on for some more minutes (the execution time for the scenario was set to 15 minutes), all of them continued. Several of them still found some objects that they had not seen before. This indicates that their situation awareness was less good than they believed it to be.

Some operators complained about the head-mounted display - after some time, it was not comfortable to wear anymore. Most operators liked the situatedness of telepresence, although some complained that they could not see the extensions of the robot, and thus felt could not maneuver well.

#### B. Future outlook

The results of the evaluation will be used to refine the requirements baseline and the use cases, e.g., the robot needs to be able to notify the operator when having collided with an object. This will eventually lead to a better performance, as the operator will have a better performance perception and can learn from his mistakes.

Furthermore, another evaluation of the methodology will be done, with refined metrics for the unit tasks and scenarios (among other things to improve the comparison), and larger numbers of end-users. In this way the data-set increases to convey systematic correlations between unit tasks and scenario operations, and the effects of individual differences. We do this by

- evaluating whether the test battery is predictive for the performance and situation awareness in a real scenario for another robot (i.e., the NIFTi robot);
- extending the evaluation mentioned above by having more participants execute the test battery tasks and the scenario;

- determining for which aspects of performance and situation awareness, the test battery task results can be used reliably as a standardization measure.

In addition, we will do further research on the general expressiveness of the unit task performances. We will especially look into for which questions the performance evaluation with unit tasks can be used and the advantages that lie in the performance of unit tasks. In particular, we are planning to apply unit test results for

- determining how much and in which way do individual operator differences play a role in the interacting with the robot and the human-robot performance;
- evaluating whether a robot is adequate for executing a particular task;
- determining whether robot-operator cooperation is clearly unsatisfactory, which might lead to either
  - determining whether an operator needs extra training in operating the robot, or
  - determining which components (hardware, software, and interaction possibilities) of a robot need to be improved.

#### ACKNOWLEDGMENT

We would like to thank the fire fighters of the city of Dortmund, Germany, and of SFO in Italy for their support. This research is supported by the EU FP7 ICT Programme, Project #247870FP7 (NIFTi), and by the Netherlands Defense UGV research program V923.

#### REFERENCES

- [1] J. van Diggelen, R. Looije, T. Mioch, M. A. Neerinx, and N. J. J. M. Smets, "A usage-centered evaluation methodology for unmanned ground vehicles," in *Proceedings of the Fifth International Conference in Computer-Human Interactions (ACHI 2012)*, Valencia, Spain, 2012.
- [2] M. A. Neerinx and J. Lindenberg, "Situated cognitive engineering for complex task environments," in *Naturalistic Decision Making and Macrocognition*, J. M. C. Schraagen, L. Militello, T. Ormerod, , and R. Lipshitz, Eds. Aldershot, UK: Ashgate, 2008.
- [3] N. J. J. M. Smets, J. M. Bradshaw, J. van Diggelen, C. M. Jonker, M. A. Neerinx, L. J. V. de Rijk, P. A. M. Senster, M. Sierhuis, and J. O. A. ten Thije, "Assessing human-agent teams for future space missions," *IEEE Intelligent Systems*, vol. 25, no. 5, pp. 46–53, September/October 2010.
- [4] A. Jacoff, E. Messina, and J. Evans, "Experiences in deploying test arenas for autonomous mobile robots," in *Proceedings of the 2001 Performance Metrics for Intelligent Systems (PerMIS) Workshop*, Mexico City, Mexico, 2001.
- [5] C. Jansen and J. B. F. van Erp, "Telepresence control of unmanned systems," in *Human-Robot Interactions in Future Military Operations*, M. Barnes and F. Jentsch, Eds. Ashgate Publishing Limited, 2010, pp. 251–270.

# Evaluation of Cognitive Effort in the Perception of Engineering Drawings as 3D Models

Florin Gîrbacia

Computer Aided Design and Virtual Reality Department  
 Transilvania University of Brasov  
 Brasov, Romania  
 garbacia@unitbv.ro

**Abstract**—In this paper, an experiment aiming at estimating the cognitive complexity of engineering drawings by measuring the reaction time and the accuracy of mentally reconstructing 3D objects from engineering drawings is presented. The performed experiment emphasizes the complexity that the engineer is facing in the product development process, shown by increased reaction time and reduced accuracy of 3D reconstructed objects. The precision and the reaction time were not considerably improved by using 3D stereoscopic viewing. Based on the results obtained from the experiment, a new class of technical drawings called Augmented Reality Technical Drawings (ARTD) is proposed. This solution enhances the visual perception by co-locating the 3D virtual object with the corresponding engineering drawing and offers the quick recognition of the object with less perceptive ambiguities.

**Keywords**—computer aided design; product development; engineering drawings; mental reconstruction.

## I. INTRODUCTION

Since the industrial age, engineering design has become an extremely demanding activity. New instruments and tools have been continuously invented in order to aid the engineer during this activity and this increase their productivity. One of the most important technologies from the last century that significantly increased the productivity in design is Computer Aided Design (CAD). Nowadays, CAD is a mature technology, without which the industrial design cannot be conceived. CAD systems have been proved to reduce the design time, costs and improve the design quality [6].

Engineering design requires sketches as well as drawings, commonly referred to as engineering drawings, taught in most engineering colleges [10]. Engineering drawings store geometrical data relating to mechanical components and assemblies, in the form of 2D planar views, either on paper or computer files. CAD systems are the main tools that help the engineer create accurately and correctly 2D drawings. CAD tools improve the productivity and the cognitive complexity by using comprehensive databases and intuitive (real-like) 3D graphic representations. Although CAD systems use 3D representation, the actual technical drawings are based just on 2D planar representations of the product. These techniques require the perception and understanding of spatial information from 2D planar

representation of a 3D object. The reconstruction problem is difficult because the 3D perception is determined from 2D planar projections and sections of the part. The user has to create a 3D image in the brain, by using 2D projections, which than is used in the manufacturing process of the respective product. In the case of a product with a high complexity, the mental effort can have a negative effect on the reaction time and the accuracy of the mentally reconstructed 3D image of the product.

Our research activities are focused on reducing the cognitive complexity that the engineer is facing. This paper presents an experimental study conducted with the purpose of analyzing the cognitive effort of using 2D engineering drawings for mentally reconstruction of 3D objects. The role of 3D stereoscopic visualization is also investigated. The paper is organized as follows: Section 2 reviews prior work, Section 3 describes the conducted experiment, Section 4 points up the results, and Section 5 presents conclusion and sketches possible directions for future work.

## II. RESEARCH BACKGROUND

The investigation of perceiving 3D object has been carried out for many years [1], [5], [9]. There is, however, a shortage of reported work that qualifies the cognitive effort of using 2D engineering drawing for mental reconstruction of 3D parts.

Hoffmann et al. [7] evaluate the perception of 3D surfaces that have been rendered by a set of lines drawn on the surface and the role of binocular disparity as a depth cue. Their results indicate that binocular and monocular mechanisms for 3D shape reconstruction from contours involve similar mechanisms.

More recently, in [8], the reaction time and the accuracy of creating brain images of 2D or 3D figures are explored. The set-up consists of performing four types of tasks: simple 2D, selective 2D, 2D-3D and 3D tasks. Their results indicate that the task of constructing the image of 3D objects in the brain has the longest reaction time. In the study were used simple and basic figures but in engineering drawings are used complex shapes.

With this in mind the objective of this research was to carry out the human factors evaluation related to mentally reconstruction of 3D parts from 2D engineering drawing via the use of a comparative assessment of monoscopic visual perception with 3D stereoscopic visual perception.

### III. EXPERIMENTAL PROTOCOL

This study tries to answer to the following research questions:

- 1) What is the accuracy and the reaction time needed to mentally reconstruct a 3D model from 2D engineering drawing?
- 2) What is the advantage of using 3D visualization for the recognition of mentally reconstructed 3D CAD models?
- 3) How engineering drawings can be improved in order to reduce the cognitive complexity?

Thus, we have devised and conducted an experiment to record and measure the reaction time and the accuracy in the brain images from several 2D engineering drawings. The results of this experiment allow us to answer the three research questions and to assess the relative impact of VR technologies on the users' performance.

#### A. Subjects

In the experiment, eleven volunteered subjects with healthy sense of vision were tested. One subject was familiar with the stimuli and with the research. The other subjects never used VR immersive stereoscopic 3D visualization for the perception of 3D CAD models until the experiment. Instead, they had extensive experience in using 2D engineering drawings and good computer skills.

#### B. Stimuli

In the conducted experiment was needed a set of sixty 3D CAD parts with medium complexity. The test parts were selected from a mechanical engineering drawing handbook and had different topology. In the first phase, each part was modeled using the legacy CATIA CAD software. In the second phase, 2D engineering drawings corresponding to CAD models were created via the drafting module of CATIA CAD software. The engineering drawings include orthographic views and for some parts sections and detailed features (Fig. 1).

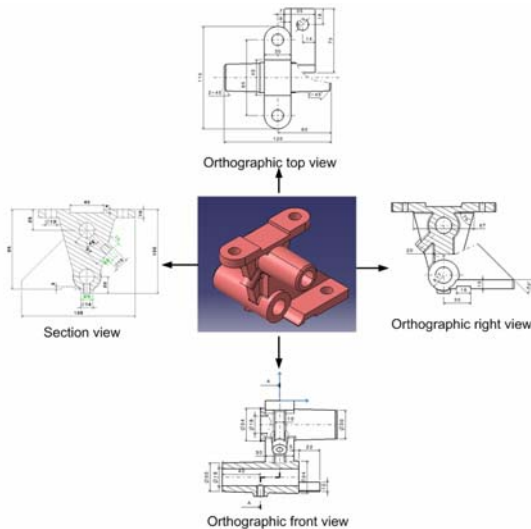


Figure 1. Components of an engineering drawing

Orthographic projections are a collection of 2D drawings that represent projections from different directions of a 3D object. Usually front, side and plan views are drawn so that a person, looking at the drawing, can reconstruct the 3D object. The sectional view is applicable to objects where the interior details are difficult to understand through the usage of dotted lines on an orthographic projection. The sectional view is obtained by cutting the object by a plane so that the hidden details are visible. Because we want to investigate the influence of stereoscopic binocular 3D vision, the parts were exported in Virtual Reality Modeling Language (VRML) neutral format and then imported in a custom developed software application.

#### C. Display technology

The 3D models were visualized using two type of devices: a universal 2D LCD display with the diagonal of 19" for desktop interface and an active stereoscopic system for 3D perception (fig. 2). For perceiving the 3D stereoscopic images, a standard solution based on desktop CRT monitors with wireless Crystal Eyes shutter glasses was used. The display renders the images at 120 Hz. For rendering the 3D environment was used the BSContact Stereo VRML visualization player integrated with a dedicated software C++ application (fig. 2).



Figure 2. The subject selecting a 3D model using 3D display active stereoscopic visualization system

#### D. Tests set-up

Before conducting the experiment, each participant had the opportunity to become familiar with the system. The users had ten minutes prior the experiment, for practicing the navigation interaction modalities in a 3D testing environment. After that they pressed the "Start" button. A random 2D drawing is initially presented and the subjects imagine a 3D model in the brain. In order to evaluate the cognitive effort of the mental reconstruction and the reaction time, for the first five parts of the experiment, the 2D drawing were presented for maxim 20 seconds. If the subjects clearly constructed in the brain the 3D model before the time passed, they click the "OK" button and step forward to the 3D environment.

Immediately after that, the subjects chose between the imaged 3D model and the other three slightly different 3D models (Fig. 3). To create a difference between the presented 3D models some features were added or removed from the part displayed in the 2D drawing (for example adding a slot to the part, or removing the ribs features). Subjects were instructed to pick with the mouse the 3D model considered to be the correct one. Each of the first five versions of 3D models was presented for maxim 25 seconds. If the user didn't make a selection, the trial is considered to be a failure. For the other five model used in the experiment, the 2D drawing was presented for maxim one minute and the 3D parts were presented for maxim 20 seconds. Half of the users first took the test using the traditional desktop system, then after a break of 10 minutes, they were asked to take the test using stereoscopic visualization system. Simultaneously, the other half of subjects first selected the correct 3D models by using stereoscopic system and then using monoscopic desktop system. There were used for the test different sets of CAD models. Each subject accomplished 20 trials. Log files recorded time stamps for the selected 3D part. This provided a task completion time (TCT) in milliseconds for each subject. The log files were used afterward for the assessment of the results. Before accomplishing the test, each participant becomes familiar with the system.

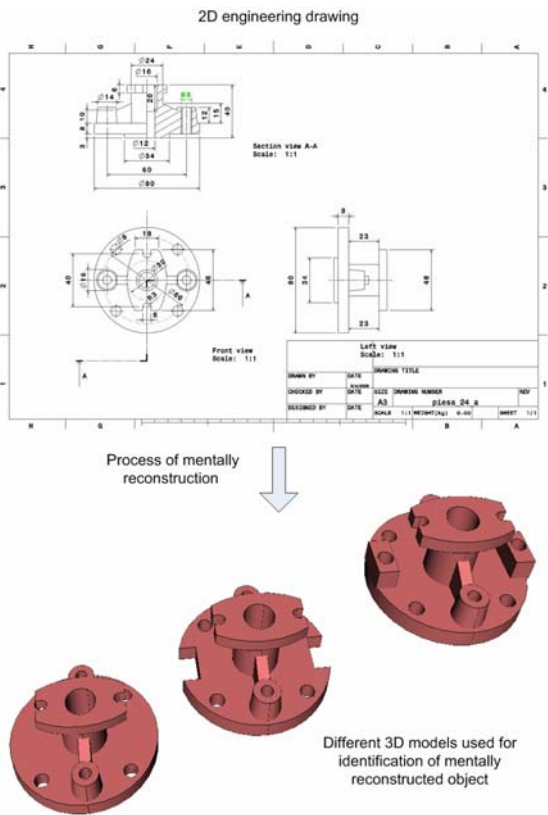


Figure 3. The process of mentally reconstruction and identification of a 3D model from a engineering drawing

#### IV. RESULTS EVALUATION

In Table 1, the mean of the correct answers from all the subjects that participated to the experiment is presented. The main reason of the significantly decreased accuracy might be the limited short term memory capacity of humans, which makes it difficult to choose between the three different 3D models, considering the 3D model that they constructed in the brain. There might occurs the situation when the mentally reconstructed 3d model is not the same with the visualized one. In this case, the used does not recognize the 3D object independently on the type of visualization (2D or 3D). When the time for visualizing the 2D drawing is increased to maximum 60 seconds, the accuracy is superior because the model could be comprehended better with the use of external working memory. Another significant result is that the stereoscopic 3D visualization does not considerable increase the recognition accuracy of imaged 3D models.

TABLE I. MEAN REACTION TIME OF SUBJECTS

Maxim time (s)	Accuracy monoscopic viewing	Accuracy stereoscopic viewing
25	43,20%	46,60%
60	56,60%	63,20%

Figure 4 presents the mean of the reaction time measured in the experiment. The analysis of these data indicates that monoscopic and stereoscopic 3D shape reconstruction involve similar methods and do not have major influence on the reaction time. The increased reaction time shows the complexity the engineer facing in the process of developing the product. The main cause might be due to the mentally manage of a large amount of information.

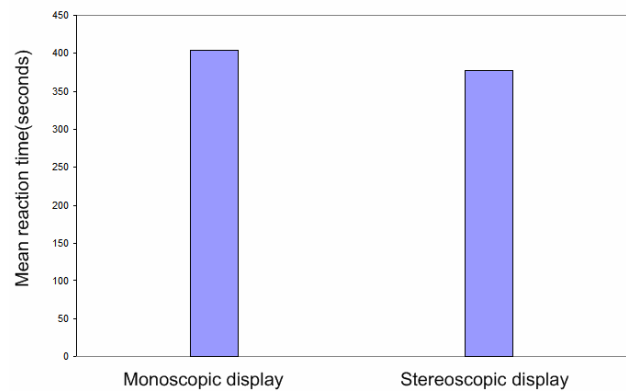


Figure 4. Mean reaction time of the subjects

#### V. CONCLUSION AND FUTURE WORK

In the process of developing the product the engineer must mentally manage a large amount of important information which leads to high cognitive complexity. Available engineering drawings do not reduce the complexity that the engineer is facing or help him handle it. The 2D engineering drawings should be easy to use and

should help the engineer manage the design-related complexity. The cognitive complexity can lead to a situation in which important design factors are not taken into consideration and this leads to failure.

From the conducted experiment, because of the reduced accuracy and the high reaction time, we can conclude that the user should visualize a three dimensional model of the part co-located with the 2D drawing.

Realistic perception of the 3D models from 2D engineering drawings plays an important role in decision making of design engineers. Recent studies [3], [4], [11] show that a system with a three dimensional representation of the model increases the performances of the users carrying out tasks which require the perception and understanding of spatial information. Unlike Virtual Reality (VR) systems, in which users are completely immersed in the virtual environment, AR users see the virtual objects and the real world co-existing in the same space (co-located). The co-location of the 3D CAD models in the real environment provides the possibility of a realist perception of the physical engineering drawing. Further work is focuses on the development of a new class of engineering drawings based on AR technologies. In Figure 5, a prototype of a developed Augmented Reality Engineering Drawing (ARTD) is shown. ARTD enhances the visual perception by co-locating the 3D virtual object with the corresponding engineering drawing and offers the quick recognition of the object with less perceptive ambiguities. This is an improvement for engineering applications where users must have a precise and direct appreciation of product shape.

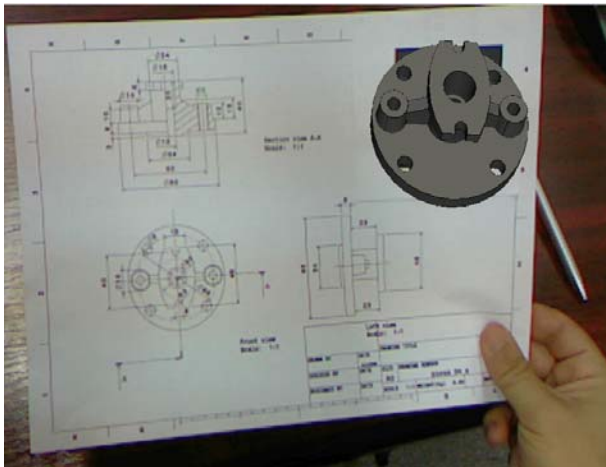


Figure 5. Augmented Reality Engineering Drawing prototype

#### ACKNOWLEDGMENT

The research activities were supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU/89/1.5/S/59323.

#### REFERENCES

- [1] D. Butler, "Predicting the perception of three-dimensional objects from the geometrical information in drawings," *Journal of Experimental Psychology: Human Perception and Performance*, Oct. 1982, vol. 8, pp. 674-692, doi: 10.1037/0096-1523.8.5.674.
- [2] D. Dori and K. Tombre, "From engineering drawings to 3D image models: are we ready now?," *Computer-Aided Design*, vol. 27, April 1995, pp. 243-254, doi:10.1016/0010-4485(95)91134-7.
- [3] P. Dunston, X. Wang, M. Billingham, and B. Hampson, "Mixed reality benefits for design perception," in *Proc. of 19th International Symposium on Automation and Robotics Construction (ISARC 2002)*, Sept. 2002, National Institute of Standards and Technology Special Publication, vol. 989, pp. 191-196.
- [4] M. Fiorentino, G. Monno, and A. E. Uva, "Tangible digital master for product lifecycle management in augmented reality," *International Journal on Interactive Design and Manufacturing*, vol. 3, May 2009, pp. 121-129, doi: 10.1007/s12008-009-0062-z.
- [5] S. Grossberg, "Cortical Dynamics of Three-Dimensional Figure-Ground Perception of Two-Dimensional Pictures," *Psychological Review*, vol. 104, Jul 1997, pp. 618-658, doi: 10.1037/0033-295X.104.3.618.
- [6] F. Garbacia, A. Beraru, and E. Butila, "Depth Perception Evaluation of 3D Virtual Prototypes Using Immersive Cave Systems," in *Proc. of Romanian National Conference of Human-Computer Interaction (RoCHI 2010)*, Matrix Rom Publisher, Sept. 2010, pp. 51-54.
- [7] C. Hoffmann, Z. Pizlo, V. Popescu, and S. Price, "Pictures Perception of surfaces from line drawings," *Displays*, vol. 28, Feb. 2007, pp. 1-7, doi:10.1016/j.displa.2006.11.001.
- [8] K. Kashihara, "Evaluation of the Cognitive Process during Mental Imaging of Two- or Three-Dimensional Figures," *Proc. of the 2009 Second International Conferences on Advances in Computer-Human Interactions (ACHI '09)*, IEEE Press, Feb. 2009, pp. 126-130, doi: 10.1109/ACHI.2009.57.
- [9] D. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artificial Intelligence*, vol. 31, March 1987, pp. 355-395, doi: 10.1016/0004-3702(87)90070-1.
- [10] L. Piegl, "Ten challenges in computer-aided design," *Computer-Aided Design*, vol. 37, April 2005, pp. 461-470, doi:10.1016/j.cad.2004.08.012.
- [11] A. E. Uva, S. Cristiano, M. Fiorentino, and G. Monno, "Distributed design review using tangible augmented technical drawings," *Computer-Aided Design*, vol. 42, May 2010, pp. 364-372, doi:10.1016/j.cad.2008.10.015.

## A Recommendation Method Based on Contents and User Feedback

So-Ryoung Kim

Department of Computer Science  
Yonsei University  
Seoul, Republic of Korea  
soryoung@cs.yonsei.ac.kr

Lae-Hyun Kim

Bionics Center, Korea  
Institute of Science and Technology,  
Seoul, Republic of Korea  
laehyunk@kist.re.kr

Sang-Min Choi

Department of Computer Science  
Yonsei University  
Seoul, Republic of Korea  
jerassi@cs.yonsei.ac.kr

Yo-Sub Han

Department of Computer Science  
Yonsei University  
Seoul, Republic of Korea  
emmous@cs.yonsei.ac.kr

**Abstract**-Nowadays, user is provided with many contents, which the previous search engines failed to find, thanks to various recommendation systems. These recommendation algorithms are usually carried out using collaborating filtering algorithm, which predicts user's preference, or contents based algorithm, which calculates on the basis of the similarity between contents. In addition to the above algorithms, many algorithms using user's context have been recently developed. Based on the previous researches, this paper proposes a new system to categorize contents information into various factors and learn user's selection. First, we divide information of items into four types and make user preference pattern using each information type. The information types can express more various user preferences and user preference pattern can calmly deal with user preference. Then, we calculate the score for recommendation using user preference pattern. That is, our system is constructed on these three modules: item analyzing module, user pattern analyzing module and recommendation score module. Lastly, we provide entire system flow to show how they work.

**Keywords**-Recommend method; Learning algorithm; User Preference; Recommendation system.

### I. INTRODUCTION

With the help of recommendation systems, people can access a wider range of contents, which was not possible with the old search engines. For example, when we listen to music or buy contents like DVD titles, books, and laptops, the recommendation systems provide a rich selection of items to aid our decisions. With the growth of the Internet society, web systems are now capable of managing and processing a large amount of information. The recommendation systems provide more intelligent information considering the user's preferences. Google search engine and Amazon recommendation systems are well-known examples.

The recommended items' properties, the user's properties, and the user's current context are the fundamentals of the recommendation systems [1]. In the case of movie recommendations, the movie properties include genre, title, and cast. These properties are used to calculate the similarity between the movies. The user

properties include the user's favorite genre, actors, and/or directors. For example, if a user likes action movies, the recommendation system will include more action movies in the recommended set. The recommendation system sets the basic direction based on items' properties and user's properties of information [2]. The user context information represents the environment the user's in, which highly influences the user's decisions [3, 4]. For instance, people prefer to watch scary movies in the summer than in the winter. The recommendation system can provide better recommendations taking the user's context information into consideration.

Many different algorithms and effective recommendation systems have used items' properties, user's properties and user's current context of information. Collaborative filtering is the most popular recommendation algorithm used in many recommendation systems [5]. This algorithm predicts the user's interests based on the user's preferences from the past. Many researchers tried to improve the algorithm using different approaches. The modified algorithms take the user's temporal context or genre preferences into consideration. This paper proposes an analysis method to categorize the information types and a learning algorithm to predict the user's item preferences based on the three information types mentioned above.

In Section II, we check the previous researches on recommendation systems using properties and learning algorithms. In Section III, we propose a new method of analyzing contents and present our recommended architecture. Then, in Section IV, we show how our system works using the system flowchart. Finally, the conclusion is given in Section V

### II. RELATED WORK

A recommendation system compares a user profile with some reference characteristics and seeks to predict the rating or preference that a user would give to an item that they had not yet considered. There are lots of recommendation algorithms. One of them is the recommendation system based on the contents properties.

This system calculates a similarity between items and suggests contents to user using this calculated similarity. Another approach is learning algorithm. The learning algorithm analyzes the user pattern when user selects contents and applies these analyzed results to the recommendation.

Our system is based on contents properties and learning algorithm for user preference.

#### A. A recommendation system based on contents properties

A recommendation system based on contents properties consists of three steps. The first step is to calculate the correlation coefficient using contents preferences. The second step is to choose neighbors of content A selected by user. Neighbors mean a group of contents that have similar rating with content A. The third step is to estimate the preference for contents based on the neighbors ratings. The detailed explanation is given in the followings.

##### - Calculating the correlation coefficient

To calculate the correlation coefficient of the contents selected by user, we use Equation (1) Pearson correlation coefficient [6, 7, 8]:

$$\rho_{xy} = \frac{\sigma_{xY}}{\cos(X,Y)} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} \times \sqrt{\sum(Y_i - \bar{Y})^2}}, \quad (1)$$

where

- X is a content.
- $\bar{X}$  is the average rating of X.
- $X_i$  denotes the rating of the  $i^{\text{th}}$  user for X.
- Y is all the other contents
- $\bar{Y}$  is the average rating of Y
- $Y_i$  denotes the rating of the  $i^{\text{th}}$  user for Y.
- $\rho_{xy}$  is the Pearson correlation coefficient between X and the other contents Y.
- *Selecting neighbors*

In this step, neighbors are chosen using the results of Equation (1). A certain correlation coefficient value close to 1 is first selected as a threshold and contents with a correlation coefficient value greater than this threshold are selected as neighbors.

##### - Predicting preferences

The final step is to predict preferences based on the ratings of neighbors. This step uses Equation (2):

$$P = \bar{X} + \frac{\sum_{y \in \text{raters}} (Y_i - Y) \rho_{xy}}{\sum_{y \in \text{raters}} |\rho_{xy}|}, \quad (2)$$

where

- $\bar{X}$  is the average rating of content X.
- $Y_i$  is the rating of the  $i^{\text{th}}$  item by other users.

- Y is the average the rating of the given content by the neighbors of X.
- $\rho_{xy}$  is the Pearson correlation coefficient between X and the other contents Y.
- *raters* are a set of contents which have ratings.
- P is the predicted value of contents for a specific user.

#### B. Learning algorithm

Machine learning is concerned with the development of algorithms allowing the machine to learn via inductive inference [9]. This inductive inference is based on observation data that represents incomplete information about statistical phenomenon. One of machine learning systems attempts to eliminate the need for human intuition in data analysis while others adopt a collaborative approach between human and machine. In our approach, we use association rules for the learning method.

##### - Association rule learning

Association rule learning is a popular and well known method for discovering interesting relations between variables in a large database [10]. Many algorithms for generating association rules were presented over time. Some well-known algorithms are Apriori, Eclat and FP-Growth, but they only do half the job, since they are algorithms for mining frequent item-sets. Another step needs to be done after to generate rules from frequent item-sets found in a database [11]. Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time.

Association rule learning divides into two separate steps. In the first step, minimum support is applied to find all frequent item sets in a database. In the second step, these frequent item sets and the minimum confidence constraint are used to form rules. Finding all frequent item sets in a database is difficult because it involves searching all possible item set.

### III. OUR APPROACH

Users, generally, selects contents on the basis of a variety of information. For example, if a user selects movie *Scream*, we assume that user likes horror genre or series. Thus, analyzing the user selections can help to predict user preference.

According to Korean Film Council (KOFIC) [12] survey, user considers genre, rating and casts when he/she chooses a movie. Table I shows the values that user considers when he/she selects movie.

We suggest a new classification method of the factors of items. We divide information of items into four types, basic information, create information, property information, related information. Then, we make user preference pattern



using each information type and calculate the score for recommendation using user preference pattern. Therefore, our system is composed of these three modules: item analyzing module, user pattern analyzing module and recommendation score module.

A. Item Analyzing Module

In item analyzing module, we first analyze properties of items. Then, we classify those analyzed factors by information types.

- **Basic information** is associated with basic information for items. Basic information includes three factors of item such as year, cast, country.
- **Create information** is associated with information of the production. Create information includes three factors of item such as creator, director, producer.
- **Property information** is associated with properties of items. Property information includes three factors of item genre, keyword, episodes.
- **Related information** is associated with information related with items. Related information includes three factors of item awards, top250, soundtrack.

B. User Pattern Analyzing Module

This module produces user preference pattern using the result of item analyzing module. User preference pattern is the configuration of individual weights on the given information type. Each information types have two patterns, short-term preference pattern and long-term preference pattern.

Short-term Preference Pattern (SPP) is the most basic pattern to define relation between items and user preferences. Preferences for items are defined by common factors between criterion item and compared item. SPP is calculated by each information type. Thus, the pattern has a total of four types, basic information pattern, create information pattern, property information pattern and related information pattern. These four patterns are used to calculate long-term pattern (LPP). Equation (3) shows how to draw the SPP.

$$SPP = \frac{|A \cap B|_i}{\sum |A \cap B|_i} \quad , (3)$$

where

- SPP is Short-term preference pattern.
- A is criterion item, B is compared item.
- *i* is the selection number of compared item.

Long-term Preference Pattern (LPP) represents learned preferences of user. It is configured about each information type and is determined by the average of the weights used in

the past and short-term preference pattern of a selected item. Equation (4) shows how to draw the LPP.

$$LPP_i = \frac{SPP_{n-1} \times LPP_{n-1}}{2} \quad , (4)$$

where

- LPP is long-term preference pattern
- *i* is the selection number of compared item.

C. Recommendation Score Module

The recommendation score module calculates the prediction score of items. It is value to determine recommendation item. Equation (5) shows how to compute the recommendation score.

$$RS_i = \sum LPP_i \times Rating_i \quad , (5)$$

where

- RS is recommendation score.
- *i* is item number.
- Rating<sub>*i*</sub> is rating on item *i*.

IV. SYSTEM FLOW

Our whole system uses values gained from SPP, LPP, and RS in Section III. Its system flow has two steps, shown in Figure 1.

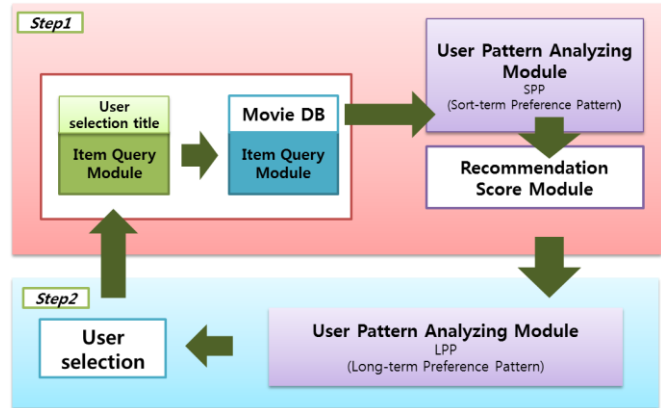


Figure 1. System flow

In Step 1, when user puts the title of item into the system, item analyzing module analyzes it based on information type. Then, analyzed data are compared with other data. The system has three small steps. First, the system analyzes the properties of compared item. Secondly, it makes short-term preference pattern. Thirdly, the system compares the compared item and input data to calculate recommendation score with the weights put on each information type. Lastly, the system provides a recommendation list for user. For example, if user inputs the title *Harry Potter*, our item query module classifies its information, such as the genre *Fantasy* and cast *Emma Watson*. After classifying information of the item, user

pattern analyzing module compares common information factors between *Harry Potter* and other movies and makes user preference pattern for recommendation. Through all these process, the system makes recommendation score and suggests a recommendation list. If the score is high in Property information type, then it recommends other *Fantasy* genre movies like *The Chronicles of Narnia*. Or if it is high on Basic information, it suggests *Emma Watson's* other movies.

In the Step 2, user pattern analyzing module operates right after user selects an item. First, it makes long term preference pattern comparing the weights of selected item and the weights in user profile. Here, the selected item becomes user input data. Ending this process, the system restarts step 1. In the *Harry Potter* case, if user selects *The Chronicles of Narnia*, then user's property information weight increases and a next recommendation list has more contents related to the genre *Fantasy* or key word *wizard*.

### V. CONCLUSION AND FUTURE WORKS

For effective recommendation, we have suggested a new classification and learning method. Using the new classification method, we have divided information of item into four types. Analyzed information types can express more various user preferences. We also have made user preference pattern. It is learned to follow up preference of user whenever he/she selects an item. Thus, our system has possibility to recommend items to user more precisely than other systems because we use more information in items and history of user pattern.

In future works, we aim to test our system using large size of database such as IMDB and to verify our algorithm. And we will also expand our algorithm using user context. For example, we can check user preference variation over time or user's feeling.

### ACKNOWLEDGEMENTS

This research was supported by the IT R&D program of MKE/IITA 2008-S-024-01 and the Basic Science Research Program through NRF funded by MEST (2010-0009168).

### REFERENCES

- [1] C. Ono, Y. Takeshima, Y. Motomura, and H. Asoh, "Context-Aware Users' Preference Models by Integrating Real and Supposed Situation Data", IEICE Trans. Inf. & vol. E91-D, no.11 November 2008
- [2] H. Asoh, C. Ono, Y. Motomura, "A Movie Recommendation Method Considering Both User's Personality and Situation", ECAI 2006, August
- [3] C. Ono, M. Kurakawa, Y. Motomura, and H. Asoh, "Implementation and evaluation of a movie recommender system considering both users' personality and situation," IPSJ J., vol. 49, no. 1, pp. 130-140, 2008
- [4] I. Zekerman and D.W. Alberecht, "Predictive statistical models for user modeling," User Modeling and User-Adapted Interaction, vol.11, no.1-2, pp.5-18, 2001.
- [5] M. Balabanovic and Y. Shoham. Feb: content-based collaborative recommendation Communications of the ACM, Vol 40, No 3, pp.66-72, 1997
- [6] D. Billsus, M. J. Pazzani, in: The 15<sup>th</sup> International Conference on Machine Learning, pp. 46-54.
- [7] B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl, in: ACM Conference on Electronic Commerce, pp. 158-167.
- [8] B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl, in: Proceedings of the 10<sup>th</sup> International World Wide Web Conference, pp. 285-295.
- [9] Bauer, M., Pohl, W. and Webb, G. (eds.): 1997, UM97 Workshop: Machine Learning for User Modeling. Online proceedings: [http://www.dfki.uni-sb.de/\\_bauer/um-ws/](http://www.dfki.uni-sb.de/_bauer/um-ws/)
- [10] Piatetsky-Shapiro, G. (1991), Discovery, analysis, and presentation of strong rules, in G. Piatetsky-Shapiro & W. J. Frawley, eds, 'Knowledge Discovery in Databases', AAAI/MIT Press, Cambridge, MA.
- [11] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.
- [12] Korea Research Team, Movie audience preference analysis, 2009.

TABLE I. CONSIDERED ELEMENTS FOR MOVIE CHOICE

	Sex		Age					
	Man	Woman	19-24age	25-29age	30-34age	35-39age	40-44age	45-49age
<b>Movie scenario</b>	86.7	93.1	86.3	93.3	88.7	95.0	88.6	87.5
<b>Other people's rating</b>	75.6	79.9	79.4	78.0	75.0	80.9	77.1	76.0
<b>Genre</b>	74.5	79.0	74.0	73.7	79.9	77.9	77.6	77.5
<b>Cast</b>	70.0	78.4	70.6	77.5	72.5	77.9	75.6	71.0
<b>Ranking</b>	71.5	68.6	68.1	70.3	64.2	70.4	73.6	74.0
Expert rating	40.8	45.7	39.7	37.8	35.3	43.2	51.2	52.5
Director	41.9	38.4	33.8	43.5	35.8	36.7	49.8	41.5
Country	32.2	33.2	31.4	35.9	27.5	32.7	35.8	33.0
Awards	28.6	29.9	25.0	20.6	23.0	29.6	39.8	38.0

Production cost	24.9	14.2	15.2	19.1	15.7	22.6	25.4	19.5
publisher	11.1	8.7	7.4	6.2	7.8	9.5	19.9	9.0

## Developing a Human Computer Interaction Course for an Information Technology Major

Cynthia Y. Lester

Department of Computer Science  
Tuskegee University  
Tuskegee, Alabama, USA  
[cylester@mytu.tuskegee.edu](mailto:cylester@mytu.tuskegee.edu)

**Abstract** – It has been reported that even in the slowing economy in the United States, that the field of technology will continue its projected growth in the job market through the year 2018. Consequently, many institutions have begun to offer degrees that specialize not just in engineering or computer science, but in information technology, as well. Degrees are offered at the associate’s, bachelor’s, master’s, and PhD levels from a variety of two-year and four-year institutions. Students who major in information technology explore management and information theory. More specifically, students focus on current technology trends and applications as it relates to the business and communication applications of computing with special emphasis on e-business, e-commerce and business computing. As a result of the merger between technology and business many of the degree programs especially at the undergraduate and graduate levels offer specific courses in human factors in information systems or human-computer interface to cover designing systems with human behavior in mind. The aim of this paper is to present the development of an undergraduate human computer interaction course for a newly developed information technology major leading to a bachelor’s degree in information technology. The paper presents the teaching pedagogy selected for the course; course description including course outcomes and topics; and, student assessments. Also discussed are challenges in the development of the course as it relates to being cross-listed for enrollment by both information technology and computer science majors.

**Keywords** – *Human computer interaction; information technology; undergraduate students.*

### I. INTRODUCTION

In the United States, the software development industry is about a \$220 billion industry which involves approximately 50,000 companies [1]. Roughly sixty percent of the software development revenue comes from software publishing or generic products, while the rest comes from customized products commissioned by a particular customer [1]. This trend influences the software development industry and the way in which newly trained technologists view the development of products and its users.

Users of these new technologies demand more from their systems now more than ever. Long gone are the days of command-line interfaces where only those “professing to be experts” in technology usage can claim ownership in the technology wars. However, now anyone who carries a cell phone, or uses a PC to type a paper, or “friends a person on Facebook” considers themselves a technology aficionado. Moreover, as e-commerce, e-business, and e-sharing become the “norm”, there is a growing need to extend the focus of software development beyond engineering and computer science to a conversation that involves the management of software and the ability to connect customers with the software.

The aim of this paper is to present a re-designed human-computer interaction (HCI) course to be offered in a newly developed information technology (IT) major at a medium-sized four-year university. The IT major is offered in the Department of Computer Science as one of two undergraduate degrees granted within the department. The original course [2], although designed to attract students from varying backgrounds, was typically taken by computer science and engineering majors, with a handful of social science majors. Students with other backgrounds, especially those in business, when asked, stated they were not attracted to the course because of the technical implications of the course and its prerequisites. Consequently, the newly re-designed course removes some of these barriers.

The paper begins by providing a brief introduction to information technology as a discipline. The following section provides a brief history of human-computer interaction and discusses the rationale for its inclusion as a required course in the IT curriculum. The next section presents the teaching pedagogy chosen for the course, the course description, and how students are assessed. Lastly, some challenges in offering the course as an interdisciplinary course for a varied audience are discussed.

### II. THE INFORMATION TECHNOLOGY CURRICULUM

Information technology can be defined as the processing, storing, and dissemination of textual, numerical, vocal, and pictorial information by a microelectronics-based

combination of computing and telecommunications [3]. The term first appeared in a 1958 article that was published in the Harvard Business Review. The authors of the article, Leavitt and Whisler commented that “the new technology does not yet have a single established name. We shall call it information technology (IT) [4].” However, since that time, information technology has become an emerging discipline of significant importance, with degrees being offered in the discipline at increasing rates.

In 2005, the ACM Special Interest Group on Information Technology Education (SIGITE) presented its final report for baccalaureate programs in Information Technology [5]. SIGITE chose to define the IT discipline as “in its broadest sense encompasses all aspects of computing technology. IT, as an academic discipline, focuses on meeting the needs of users within an organizational and societal context through the selection, creation, application, integration and administration of computing technologies [5].”

The final report presented the IT Body of Knowledge, learning outcomes, the IT core, IT electives, curriculum models, and course descriptions. It was decided by the group that the aim of an IT program would “provide IT graduates with the skills and knowledge to take on appropriate professional positions in Information Technology upon graduation and grow into leadership positions or pursue research or graduate studies in the field [4].” More specifically, it was decided that within five years of graduation a student must be able to [5]:

- Explain and apply appropriate information technologies and employ appropriate methodologies to help an individual or organization achieve its goals and objectives
- Manage the information technology resources of an individual or organization
- Anticipate the changing direction of information technology and evaluate and communicate the likely utility of new technologies to an individual or organization
- Understand and for some to contribute to the scientific, mathematical and theoretical foundations on which information technologies are built
- Live and work as a contributing, well-rounded member of society

It was decided that in order to meet these goals that the IT Body of Knowledge would include the areas of study, identified and presented in Table I [5].

The next section presents a brief history of HCI. Also, stated in this section is the rationale for the inclusion of HCI as a required as part of the IT curriculum.

TABLE I. IT BODY OF KNOWLEDGE

Area of Study
Information Technology Fundamentals
Human Computer Interaction
Information Assurance and Security
Management
Integrative Programming & Technologies
Networking
Programming Fundamentals
Platform Technologies
Systems Administration and Maintenance
System Integration & Architecture
Social and Professional Issues
Web

### III. HUMAN COMPUTER INTERACTION: A BRIEF HISTORY

Human-computer interaction has been defined in various ways. Some definitions suggest that it is concerned with how people use computers so that they can meet users’ needs, while other researchers define HCI as a field that is concerned with researching and designing computer-based systems for people [6], [7]. Still other researchers define HCI as a discipline that involves the design, implementation and evaluation of interactive computing systems for human use and with the study of major phenomena surrounding them [8]. However, no matter what definition is chosen to define HCI, the concept that all these definitions have in common is the idea of the technological system interacting with users in a seamless manner to meet users’ needs.

Human-computer interaction has its roots embedded in the systematic study of human performance [9]. It has been stated that World War II provided the impetus for studying the interaction between machines and humans, as the military for each side attempted to manufacture more effective weapons systems [10]. These research activities led to the development of the field of Ergonomics and in 1948 the Ergonomics Research Society of England brought together researchers of varied backgrounds who had an interest in the design of equipment for human use [8]. After the World War, the focus of concern expanded to include worker safety, where more research led to the development of an additional field, human factors [11].

With the advances in technology, the terms, ergonomics and human factors are now often used interchangeably. More specifically, the terms as now defined by the International Ergonomics Association, the world’s leading organization on ergonomics and human factors defines the similar terms as, “the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and other methods to design in order to optimize human well-being and overall system performance [12].” These disciplines are noted to have provided the basis for human-computer interaction as both are concerned with user performance of any system, whether machine or

computer [8]. Consequently, HCI serves as an important area in information technology and the IT Body of Knowledge.

The Department of Computer Science at Tuskegee University developed its IT curriculum based on the SIGITE report and included as part of its core courses, human-computer interaction [5]. The rationale for the inclusion of HCI in the curriculum was two-fold. The first reason to include HCI in the curriculum was to provide students with an opportunity to gain a knowledge base in an interdisciplinary field as information technology is also multi-disciplinary. The other reason was to provide students with an opportunity to gain training in the humanistic side of technology where the focus is not only on the system but on the user as well.

The next section presents an overview of course development. Included in this section is the teaching pedagogy chosen, course description, and student assessment.

#### IV. TEACHING PEDAGOGY

##### A. Teaching style

According to Grasha, there are four styles of teaching, which encompass the following [13]:

- Formal authority – instructor-centered approach where the instructor provides and controls the flow of content for the course
- Demonstrator/personal model – instructor-centered approach where the instructor demonstrates the skills that students are expected to learn
- Facilitator – student-centered approach where the instructor acts as a facilitator and the responsibility is placed on the student to achieve results for various tasks
- Delegator – student-centered approach where the instructor delegates and places the control and the responsibility for learning on the students and/or groups of students

The facilitator teaching style was chosen because it is a student-centered approach which shifts the focus of activity from the teacher to the learners. This method includes active learning, collaborative learning and inductive teaching and learning [13]. The facilitator teaching style has been stated to work best for students who are comfortable with independent learning and who can actively participate and collaborate with other students [14]. In particular, this approach was chosen because in education literature, the method has been shown to increase students' motivation to learn, to lead to a greater retention of knowledge, and to positively impact attitudes toward the subject material being taught [15], [16], [17]. Additionally, the method places a strong emphasis on collaborative learning.

##### B. Collaborative learning

Students learn best when they are actively involved and engaged in the learning process. In educational environments, study groups are often formed to gain better insight on course topics through collaborative efforts. Collaborative learning is defined as the grouping and/or pairing of students for the purpose of achieving an academic goal [18]. Davis reported that regardless of the subject matter, students working in small groups tend to learn more of what is taught and retain it longer, than when the same content is presented in other more traditional instructional formats [19].

Supporters of collaborative learning suggest that the active exchange of ideas within small groups not only increases interest among group participants, but also helps to improve critical thinking skills. The shared learning environment allows students to engage in discussion, take responsibility for their own learning, hence becoming critical thinkers [18]. Students are responsible for their own learning as well as the learning of others. Research has shown that collaborative learning encourages the use of high-level cognitive strategies, critical thinking, and positive attitudes toward learning [20]. Further, it has been suggested that collaborative learning has a positive influence on student academic performance [16].

#### V. COURSE DESCRIPTION

The description of the course, *CSCI 440 – Human Computer Interaction*, which is set to be offered during the 2012-2013 academic year, is to provide students with an introduction to human-computer interaction and to expose them to current research topics within the field [21]. The prerequisites for the course include successful completion of *CSCI 220 – Programming II* and *IT 325 – Web Systems Design* or *CSCI 230 – Data Structures*. The prerequisites were chosen to ensure that students had some experience with structured programming and scripting languages; and, that they had completed many of the general university requirements, where some of the concepts from these courses would be used in the HCI course.

The course outcomes describe the specific knowledge and skills that students are expected to acquire. The course outcomes for *CSCI 440* include the following, and at the end of the course a student should be able to:

- Clearly state the multi-disciplinary nature of HCI and its origin.
- Identify the different areas of study within and current research topics related to the HCI discipline.
- Identify the basic psychological and physiological attributes of computer users.
- Describe and identify the components and devices of computer systems.

- Describe the fundamentals of the HCI design process.
- Explain how technologies help organizations/individuals meet their goals.

To meet the objectives of the course outcomes, the content of the course included:

- Introduction to HCI
- The Human Component of HCI
- The Computer Component
- Interaction Basics
- The Design Process
- Evaluative Techniques
- Current Topics in HCI

Table II is an outline of the topics covered during the sixteen week semester [21].

TABLE II. – COURSE SCHEDULE

Week	Topic
1	Introduction
2	Historical Perspective of HCI
3	The Human
4	The Human
5	The Computer
6	The Computer
7	The Interaction
8	The Interaction
9	Spring Break – No Class
10	Interaction design basics
11	Design rules
12	Evaluation techniques
13	Evaluation techniques
14	Universal design
15	Universal design
16	Putting it all together

Learning resources may be of the following two broad types: (a) lecture materials; and (b) development or modification of class projects and descriptions. Lecture materials introduce specific concepts, terminology, and techniques to the students in the classroom. Class projects are often used to reinforce student learning. The next section presents a description of the class projects, the paper in special topics, and the in-class debates on case studies, which are all used to assess student learning. Also, presented is how each assessment has been designed to reinforce the theoretical concepts taught in class.

## VI. STUDENT ASSESSMENT

### A. Class Projects

Three class projects were re-designed. Each project was named for a popular television show. Students are required to complete each project according to the stated specifications and then provide a written report and to

present the result of the project during the class hour. The names of the projects are:

- Designers’ Challenge
- Design on a Dime
- America’s Next Top Model

This section discusses the projects in detail.

#### 1) Designers’ Challenge

The purpose of *Designers’ Challenge* is to allow students to use their creativity to redesign an indirect input device. The specifications for this project include developing a prototype or mock-up of a hand-held device that could be used by *United Parcel Service of America, Inc.* (UPS). The following fictional scenario is provided for review by the students:

Managers at UPS have noticed that it takes a delivery person approximately 10 minutes to deliver a package, which includes having the receiver sign for the package. Managers have also noticed that it is the signing for the package by receivers that takes the longest amount of time and results in the largest number of errors. Often, customers want to chat with the UPS employee and forget to sign for the package, the UPS employee forgets to charge the device, or the UPS employee injures his or her hand when handling the device as well as the package. Therefore, to minimize error rate and to improve performance, UPS has enlisted a company to redesign their current input device receiving packages, which is an electronic signature board and pen, to a new and improved device using the latest technology for the UPS employee to use to capture the signature of the receiver of the package.

Also provided are the following fictional general requirements:

The device cannot be any larger than one 8½ X 11” sheet of paper and no smaller than the size of regular-sized PDA. The device must not weigh more than one (1) pound. Anything else would be too heavy for the UPS employee and the receiver of the package to use. The device may use text entry or a positioning, pointing, or drawing device.

#### 2) Design on a Dime

The purpose of the *Design on a Dime* exercise is to allow students to design the interface to a device. Students are required to supply the following:

- A persona that describes the core user group
- A scenario that describes an example of how the ordering tool will be used
- A network diagram that illustrates the main screens or states of the ordering tool
- A functional prototype of the ordering tool

Students are given the following fictional scenario:

A new Apple store is coming to town. Apple expects the store to be wildly popular since it will cater to the more

than 3000+ students, faculty, and staff at Tuskegee University. To get the store up and running it has decided to allow students in the *CSCI-440* class to design an interface for its new inventory tool.

### 3) America's Next Top Model

The purpose of this project is to allow students an opportunity to use the experimental evaluation techniques discussed in class to analyze and assess user interaction with software. Students are required to select two popular search engines currently being used and to conduct an experimental evaluation to determine if the products are successful according to today's users. The method that students are encouraged to use, include the following:

- Choose at least six but no more than ten of their peers to participate in the study
- Develop a series of tasks that each person is to perform
- State two hypotheses that can be tested
- Use statistical tests to display results
- Use descriptive statistics to make inferences about the population

The written report must include the following:

- A statement of the problem (purpose of the study)
- An introduction to the search engines
  - Important features of each
  - Illustration of each
- A description of the evaluate technique used, including the stated hypotheses
- Results
  - Description of the population
  - Descriptive statistics describing results (i.e., charts, graphs)
- Discussion of results and concluding thoughts
- An appendix containing the hard copies of the end user survey

### B. Collaborative Learning

The premise of the collaborative learning was to expose students to current case studies in the field and then allow them to discuss the pros and cons of the case studies. In the field of software security it has been stated that case studies provide an opportunity to review the practical aspects of software security [17]. Therefore, this concept was applied to HCI as it relates to current IT trends.

The directions stated that the student should do the following:

- Analyze the situation
- Use analogies and similar cases where possible
- Identify possible risks or consequences
- Discuss how the new technology changes the situation, if relevant

- Discuss the advantages or problems resulting from using the "new and now" technology and compare it to the "old and then" use of technology
- Present your opinion of the situation

Additional directions include that the presentation of each student team should be no less than fifty minutes in length which covers the entire class period. Each team is also required to submit a typed-written paper, not to exceed one (1) page, which includes the aforementioned points. The paper must include at least three (3) references from which the information was gathered.

### C. Special Topics Paper

Students are to select any current topic in the field of HCI to research, which is not presently covered in class. Students are required to write a research paper on the topic and to present the paper in class. The paper has to be type-written and between eight and ten pages in length. Additionally, students are required to follow either the *IEEE Computer Society Style Guide* or *The Publication Manual of the American Psychological Association*. A list of research papers is made available to the students along with suggested topics for presentation.

The aim of the technical-writing component allows students to practice their technical writing skills as well it reinforces the concepts of writing a technical literature review. Lastly the special topics paper also allows students to choose a topic of their interest and present the concepts not previously explored in class.

## VII. CHALLENGES

The author encountered several challenges in re-designing the course. The first challenge that the author faced was course enrollment and constituency [22]. Although originally designed from a multi-disciplinary perspective, the overwhelming majority of the students enrolled in the course were computer science and engineering majors with a very small sprinkling of social science majors. The original course was designed to take advantage of students' skill set in programming, technical writing, and math (as it relates to statistical analysis). Therefore, to meet the needs of the needs of the original constituency group and the new IT majors, it was decided that the course would still be taken by students at the senior level, but the prerequisites would change to include those with programming experience in structured and or scripting languages. It was further decided that keeping the course as a senior level course would ensure that all had completed general education university requirements, ensuring the multi-disciplinary scope of the course.

The next challenge the author faced was deciding on the content for the course projects. The author wanted to ensure that information technology students who would be well-grounded in application development tools and web systems



design would be comfortable alongside their computer science counterparts who would be well-grounded in theory and programming and vice-versa. Therefore, it was decided that the projects would remain multi-disciplinary in nature, but would also encompass business principles.

Lastly, the author wanted to ensure that students left the course understanding the current topics in the ever-changing field of technology. Therefore, it was decided that the best way to engage students in this area is through the current topics paper and through the collaborative learning exercises.

### VIII. CONCLUSION

The aim of this paper was to present a re-designed undergraduate human computer interaction course for a newly developed information technology major leading to a bachelor's degree in information technology. The paper provided a high level overview of the teaching pedagogy selected for the course, a description of the course and the assessment methods chosen to examine student learning. Also, presented are certain challenges the authored encountered in further revising the course for an even more multi-disciplinary audience.

As technology increases in complexity and users demand more from their technological systems, it is imperative that academia continue its dominance in training the technologists of the future. As educators, we must provide students with more options to gain entrance into the technology race besides the traditional engineering and computer science tracks. The broad discipline of Information Technology allows students to gain an understanding of how to meet the needs of the diverse user by merging the technical side of computing with its humanistic side. From ergonomics, to human factors, to the human computer interface and beyond, the influence of humans on technology is persistent but also ever-changing creating a dynamic interplay somewhere between humans, computers, and their interaction.

### REFERENCES

- [1] First Research Industry Profile. <http://www.csb.uncw.edu/people/farinellaj/classes/FIN430/Handouts/Industry%20Profile%20for%20case%202.pdf> (Accessed on September 6, 2011).
- [2] C. Lester, C. "Advancing the Multidisciplinary Nature of Human Computer Interaction in a Newly Developed Undergraduate Course." (2008). *Proceedings of the International Conference on Advances in Computer-Human Interaction*. IEEE Computer Society Press.
- [3] D. Longley and M. Shain. (1985). *Dictionary of Information Technology* (2 ed.), Macmillan Press, p. 164, ISBN 0-333-37260-3
- [4] Information Technology. (1989). *Oxford English Dictionary* (2 ed.), Oxford University Press. (Accessed August 4, 2011).
- [5] SIGITE Curriculum Committee Special Interest Group on Information Technology Education of the ACM (Association for Computing Machinery) (2005). *Information Technology Curricula*. [http://www.acm.org/education/curric\\_vols/IT\\_October\\_2005.pdf](http://www.acm.org/education/curric_vols/IT_October_2005.pdf) (Accessed on August 31, 2011).
- [6] D. Benyon; G. Davies; L. Keller; J. Preece & Y. Rogers. (1998). *A Guide to Usability*, Addison Wesley, 0-201-6278-X, Reading, MA.
- [7] H. Sharpe; Y. Rogers & J. Preece. (2007). *Interaction design: beyond human-computer interaction 2<sup>nd</sup> ed.* John Wiley & Sons Ltd, 978-0-470-01866-8, England.
- [8] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland and T. Care. (1994). *Human-Computer Interaction*. Addison Wesley, 0-201-62769-8, Reading, MA.
- [9] A. Dix, J. Finlay, G.B. Abowd and R. Beale. (2004). *Human-Computer Interaction*. Prentice Hall, 0130-461091, Boston, MA.
- [10] L. Wilson and I. Benbasat. Designing an electronic commerce Research and Applications, Volume 2, Issue 3, Autumn 2003, Pages 240-253, ISSN 1567-4223, 10.1016/S1567-4223(03)00026-7.
- [11] FAQ for Interface Design, Human Factors & Ergonomics. <http://www.usernomics.com/ergonomics-faq.html> (Accessed September 15, 2011).
- [12] Human Factors and Ergonomics Society <http://hfes.org/web/AboutHFES/about.html> (Accessed on September 2, 2011).
- [13] A.F. Grasha. (1994). "A matter of style: The teacher as expert, formal authority, personal model, facilitator, and delegator." *College Teaching*. 42:142-149.
- [14] R.M. Felder and R. Brent. (1996). "Navigating the Bumpy Road to Student-Centered Instruction." *College Teaching*. 44:43-47.
- [15] C.C. Bonwell and J.A. Eison. (1991). "Active learning: Creating excitement in the classroom." *ASHE-ERIC Higher Education Report No. 1*. Washington, DC: George Washington University.
- [16] R.T. Johnson and D.W. Johnson. "An Overview of collaborative learning." *Creativity and Collaborative Learning*; Baltimore: Brookes Press. 1994. [Electronic Version].<http://www.cooperation.org/pages/overviewpaper.html> (Accessed on August 31, 2006).
- [17] C. Meyers and T.B. Jones. (1993). *Promoting active learning: Strategies for the college classroom*. San Francisco: Jossey Bass.
- [18] A. Gokhale. (1995). "Collaborative learning enhances critical thinking." *Journal of Technology Education* 7, no. 1.
- [19] B.G. Davis. (1993). *Tools for Teaching*. San Francisco: Jossey-Bass Publishers.
- [20] S. Wang and S. Lin. (2006). "The effects of group composition of self-efficacy and collective efficacy on computer-supported collaborative learning." *Computer and Human Behavior*.
- [21] C. Lester. (2011). *CSCI 440 – Human Computer Interaction*. Course Syllabus.

## Physical Instructional Support System Using Virtual Avatars

Tomoaki Ogawa

Department of Computer and Information Engineering,  
Nippon Institute of Technology 4-1 gakuendai,  
Miyashiro-cho, Minamisaitama-gun, Saitama 345-8501  
Japan  
c1065201@cstu.nit.ac.jp

Yasushi Kambayashi

Department of Computer and Information Engineering,  
Nippon Institute of Technology 4-1 gakuendai,  
Miyashiro-cho, Minamisaitama-gun, Saitama 345-8501  
Japan  
yasushi@nit.ac.jp

**Abstract**—Certain sports such as martial arts and dance have sets of good typical motion types. These motion types were abstracted from the physical movements of excellent practitioners. They are devised for instruction purpose, and are optimal movements of the target sports. It is extremely important for learners to learn these typical motion types. Traditionally the transfer of such typical motion types is done by in-person instructions. Therefore it is considered that it is not suitable for distance learning environment. We have developed a support system to convey this motion types by way of communication networks. In this paper, we propose a real-time physical instructional support system. The instructor and the learners communicate with each other by the virtual humanoid 3D-CG avatars through the Internet. By using this system, it is possible for the instructor to demonstrate his motions, and for learners to obtain the instructor's movements at distant places. The obtained information is beyond simple camera images. Because the system provides the three-dimensional perspectives and it superimposes the instructor's movements on the learner's avatars as well as provides a means of real-time direct communication.

**Keywords**—education; sport; Kinect; MMDAgentt; remote instruction

### I. INTRODUCTION

Certain sports such as martial arts and dance have good sets of typical motion types. These motion types were abstracted from the physical movement of excellent practitioners. They are devised for instructional purpose, and are optimal movements of the target sports. It is extremely important for learners to learn these typical motion types.

Traditionally learning these motions is done by imitation learning. In imitation learning, the learner mimics the motions of the instructor. Even imitation learning, not-in-person learning is possible. Various methods for transferring physical movements without direct instructions are developed. Self-learning through books and recorded video viewing are two popular ways to learn. None of them are as effective as direct-learning through actual experiences with instructors, though.

In order to mitigate the problems of self-learning, Chua et al developed a distance learning system that feedbacks the learner's movements to him in order to highlight differences from the recorded instructor's movements [1]. With the distance learning system, the user can observe the

instructor's demonstrations as well as his own recorded movements as many times as he wants. When the learner encounters some confusing situation, however, he has no way to communicate with the instructor to get immediate advices. It is also hard for the learner to understand the instructor's intention behind a certain physical motion.

We have devised a way of solving these problems. In this paper, we propose a real-time physical instructional support system. The instructor and learners communicate with each other by the virtual humanoid 3D-CG avatars through the Internet. By using this system, it is possible for the instructor to demonstrate his motions, and for learners to obtain the instructor's movements at distant places. The obtained information is beyond simple camera images. Because the system provides three-dimensional perspectives and it superimposes the instructor's movements on the learner's avatars as well as provides means of real-time direct communication.

In this system, the virtual avatars of the instructor and the learners share the same virtual space. Both the instructor and the learners exercise their performances in front of the sensor called Kinect. Kinect is a Microsoft product and widely available [2]. Since Kinect tracks the performers' movements in real-time, the virtual avatars in the shared virtual space move precisely as the owner of the avatar. Since Kinect features an RGB camera and a depth sensor, it can provide the depth values of the scene in real-time. Using these depth values, the system constructs the humanoid 3D avatars in the virtual space. Since each virtual avatar maintains 3D information from Kinect, the system can display the movements of the avatar from any angles by rotating the virtual camera.

Also, since the avatars of the instructor and the learners share the same virtual space, this system can express the differences between the model movement and the learners' not-so-good real movements directly. Those differences are hard to express in verbal description and camera images. In addition, the system allows resizing the avatars of learners and instructor so that they have the same size. This can compensate for the difference in physique between the two. In addition to these features, we have integrated a gesture and voice recognition and speech synthesis subsystems into our physical instructional support system so that the instructor and the learners can verbally communicate with

each other as well as they can control the virtual camera and their avatars.

Figure 1 shows the overview of the physical instructional support system we are currently constructing. In this figure, the Kinect sensor obtains 3D motions of the instructor and the learners. The local PC connected to the Kinect extracts the skeleton information of the data sent from the Kinect, and transmits it to the server. There is one central server that provides communications between local PC's so that all the participating local PC's share one virtual space where all the avatars that represent the instructor and the learners reside. The local PC constructs avatars from the skeleton information sent from the server. Each avatar corresponds to the skeleton information sent from each remote PC, which manipulates the avatar so that each avatar reflects its owner's movements in the virtual space.

The rest of this paper is organized as follows. In the next section, we describe the related works. In section III, we describe the leaning process of physical body movement. In section IV we describe the physical instructional support system we are constructing. We conclude our discussion in section V.

## II. RELATED RESEARCH

There are quite a few research works for distant learning systems that support obtaining certain physical motion types. The most notable self-learning system that is closely related to our system was developed by Chua et al.

Chua et al developed a self-learning system that uses a virtual avatar in order to learn Tai Chi. [1]. The user learns typical motions by observing instructor's pre-recorded demonstration. Then the user tries to imitate the movements. The system gives feedback to the learner's movement by using avatar and superimposes it on the instructor's model movement so that it highlights the differences between them. The learner is supposed to correct his movement by observing the difference. Unfortunately, it lacks direct communication between the instructor and the learners. It confines its effectiveness in self-study environment. Unlike their work, we have integrated a gesture and voice recognition and speech synthesis subsystems into our physical instructional support system so that the instructor and the learners can verbally communicate each other as well as they can control the virtual camera and their avatars.

Usui et al developed a dance entertainment system that consists of a recording studio and a mobile movie viewer [3]. The user can mimic a dance movie and record his movement, and then review his dance technique with other dancers' recorded movements on the mobile viewer. This system provides a convenient way of comparison learning by using a handy mobile viewer.

Honjou et al developed a self-learning system for golf swings. In the system, the user mounts a semi-transparent head-mounted display (HMD) that shows the instructor's swing concurrently when the user swing his club [4]. These studies pay attention to how to show the instructor's motion, at the same time the user performs his motion so that the user can recognize the difference between the instructor's model movement and his own movement. They all lack, however,

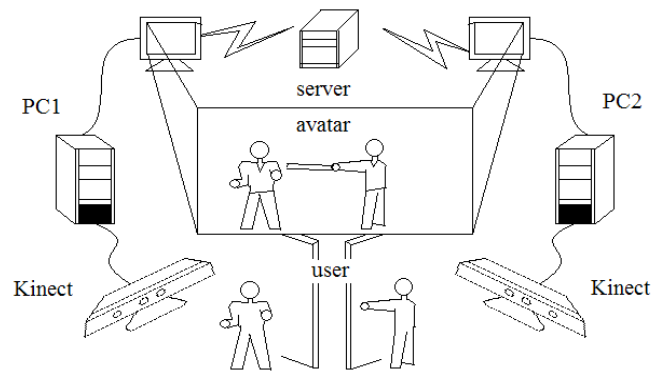


Figure 1. System overview

interactive features so that the user has no way to contact the instructors as he practices.

For real-time interaction, Nakazawa et al developed a high-definition HyperMirror system utilizing satellite communication for interactive distance learning [5]. The system provides extremely good videoconference experiences by using the real image of the attenders so that all the participants can feel as if they were in the same room. The video image, however, is hard to manipulate so that the system cannot rotate the image of people. The user cannot observe counter-part's image at different angles. Even though providing cordial feeling to the participants as if they are sharing the same space, video image is not ideal for physical instructions. In addition, Morikawa et al have reported that there exists some aversion of self-image [6].

Kimura et al studied how to display visual information for the remote learning of various body movements over network. In the system, the instructor and the learner equipped various sensors on the joint parts of the arm, and they can view the movement of the arm in the HMD screen [7]. The instructor and the learner see the virtual avatar in the shape of the arm, so that the instructor's arm position can be superimposed over the learner's. The instructor can correct the learner's movement by verbal instruction over IP telephone. They have succeeded in constructing an instructional system, but it is limited to only arm position due to the requirement of too many sensors. Our system requires no attached sensors so that the system should be able to adapt for whole movements.

So far we are not aware of any support system of interactive physical movement learning which uses the speech and gesture recognition so that the instructors and the learners can communicate as naturally as real face-to-face instruction. Such a system is what we are presenting in this paper.

## III. LEARNING PROCESS OF MODEL MOVEMENTS

In this section we discuss the learning process of certain sports that have typical physical movements. We assume that the instructor and the learners pursue the following process to teach and learn the typical movements.

1. Presentation: the instructor demonstrates his model motions to the learner.

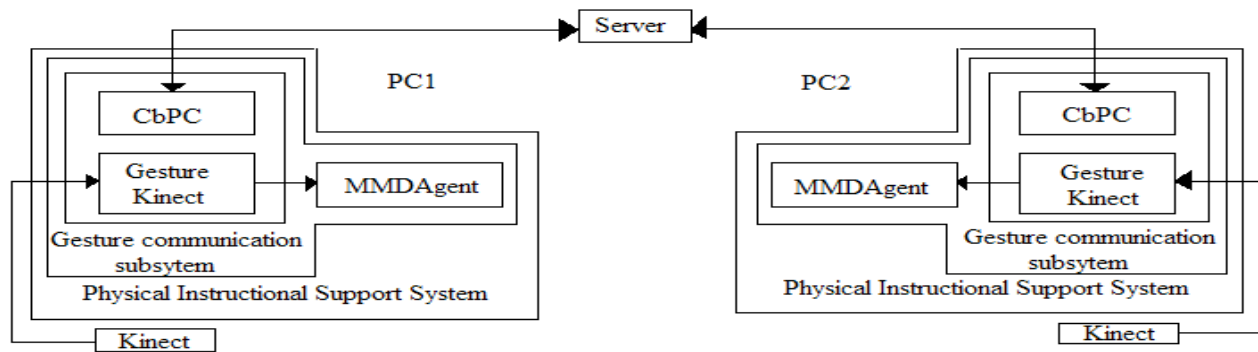


Figure 2. System configuration

2. Comprehension: the learners observe the instructor's demonstration and try to comprehend how to imitate the model motions.
3. Practice: the learner imitates the motions based on his comprehension.
4. Assessment: the instructor assesses the movements of the learner who mimics his demonstration. The instructor makes corrections on the learner's exercises with gestures and verbal advices.
5. Confirmation: the learner repeats his exercises based on his comprehension to confirm the movements and understanding of the intention of the instructor.

The most important part of the learning process is the assessment of learner's physical motions. In this assessment, the instructor measures the progress of the learner's understanding, and gives effective instructions to correct the learner's movements.

For effective instruction, the instructional support system put the first priority to the smooth sharing of information and understanding between the instructor and the learners. Our instructional support system satisfies this requirement through smooth communication between them based on the virtual avatars.

#### IV. INSTRUCTION-SUPPORT-SYSTEM

##### A. Functional requirements and solutions

In order to construct the physical instructional support system that endorses the learning process we have just described in the previous section, we set the following functional requirements and solutions.

- 1) *The system must keep tracking of the performer's movement and reflects it to the corresponding avatar.*
- 2) *The system must recognize the speech and gesture of the performer.*
- 3) *The system must provide real-time interactions between performers in distant places.*

We discuss them some details below:

##### 1) Tracking control of a virtual avatar

The system keeps tracking of the performer's movement and reproduces the instructor's and learners' movements on the virtual avatars. Each performer, either

the instructor or the learner, has its own avatar. All the avatars share the same virtual space, and all the participants at various distant places can observe all the avatars in that virtual space. In other words, they see the same virtual space from their own remote display terminals.

In order to achieve this requirement, we utilize a sensor called Kinect to keep tracking of motion of the instructors and the learners. The instructors and the learners are displaying their physical movements in front of the Kinect and the system reflects the motions on the virtual avatars in the same virtual space. We use MMDAgent for modeling of the virtual avatars [8]. MMDAgent is a Tool Kit that constructs humanoid 3D models and allows the models to communicate with the human users. The Speech Processing Laboratory of Nagoya Institute of Technology has developed MMDAgent, and the institute provides it freely. We construct the avatars by using MMDAgent.

##### 2) Speech recognition and gesture recognition

The system must rotate and move the virtual camera as well as the avatars. The system must provide a means to get the attention of instructor. The instructor and the learners should be able to verbally communicate with the system and instruct the system to rotate the virtual camera as well as the avatars. Also the learners should be able to verbally request the system to get attention of the instructor. The communication between the human users, i.e. the instructor and the learners, and the system are done through speech recognition subsystem so that they can concentrate their physical movement practices.

In order to achieve this requirement, we utilize MMDAgent for speech recognition. Also, we have integrated a gesture recognition subsystem into MMDAgent so that the users can communicate with the system via not only voice but also gestures. Figure 2 shows the system configuration. A detailed explanation about the gesture communication subsystem is given in the next subsection.

##### 3) Distant communication

The system can display avatars in distant places so that all the participants in distant places can see the same virtual avatars at the same time. The participants should

be able to perform voice-chat. The voice chat between users and the verbal commands to the system are differentiated by means of the “mode” so that the users can give commands to the system as well as can talk to each other.

In order to achieve this specification, the system has been connected with UDP in socket communication to realize high-speed connection. The gesture communication subsystem system has a means to switch between voice-chat and speech recognition for commanding the system.

**B. Gesture communication subsystem**

We have developed a gesture communication system to amend MMDAgent. This subsystem recognizes the gestures of human performers that are sent from the Kinect. The different gestures indicate whether the user wants to command the system, such as to rotate the virtual cameras and the avatars or the user wants to communicate to one of the users in a distant place. This subsystem consists of the following two parts (see Figure 2).

Part 1, Gesture Kinect, is in charge of the gesture recognition that obtains the skeleton information sent from the Kinect. The Kinect senses 3D motion of the user and constructs skeleton information as shown in Figure 3. Gesture Kinect recognizes certain gestures of the user and sends commands to MMDAgent to control the virtual camera and avatars. In order to point precise position of learner’s body, the instructor can create and use a virtual tact through gesture command. (Figure 1 shows the avatar of the instructor uses the tact to pint the learner’s body.) Table I shows typical gesture commands.

Part 2, Communications between the PC (CbPC), controls the voice-chat and transmitting the skeleton information for avatars.

**V. CONCLUSION AND DISCUSSION**

In this paper, we proposed a real-time physical instructional support system through virtual avatars. The instructor and learners who are in remote places can share the same virtual space by their avatars. This system can transmit 3D movements that are unable to be realized by the video image. Thus the instructor can give the learners

precise guidance. The instructor can point the precise position of a certain learner, and show the model movement by superimposing his avatar. Since the instructor and the learners share the same virtual space, it is possible for them to keep mutual understanding. As performing in front of Kinet sensor, the need for HMD and other sensors attached to the bodies is completely eliminated. Also the speech and gesture recognition subsystem eliminates the use of keyboard and mouse, so that they can perform, demonstrate, and learn in more natural and effective environment than all the previous similar systems. In addition, communication with avatars drastically decreases the amount of transmission data because only skeleton information just enough to construct avatars’ movement is sent and received between distant places.

**REFERENCES**

- [1] P. T. Chua, R. Crivella, B. Daly, N. Hu, R. Schaaf, D. Ventura, T. Camill, J. Hodgins, and R Pausch, “Training for physical tasks in virtual reality environments: Tai chi,” Proc. IEEE Virtual Reality. IEEE Press, 2003, pp. 87-94.
- [2] [http:// research.microsoft.com/ en-us/ um/ redmond/ projects/ kinetsdk/ default.aspx](http://research.microsoft.com/en-us/um/redmond/projects/kinetsdk/default.aspx), last access: Dec. 6, 2011.
- [3] J. Usui, H. Hatayama, T. Sato, Y. Furuoka, and N. Okuda, “Paravie: dance entertainment system for everyone to express oneself with movement,” Proc. 2006 ACM SIGCHI International Conf. Advances in Computer Entertainment Technology, 2006, pp. 14-16.
- [4] N. Honjou, T. Isaka, T. Mitsuda, and S. Kawamura, “Proposal of Method of Sports Skill Learning using HMD,” Trans. Virtual Reality Society of Japan, Vol. 10, No.1, 2005, pp. 63-70.
- [5] A.Nakazawa T, Okubayashi, H Mori, T Maesako, O Morikawa, M Nakao, N Tomii, T Sato, T Kawasugi, and G Hashimoto, “Applying High-Definition “HyperMirror” to Distance Learning Utilizing “KIZUNA” ”, 27th International Symp. Space Technology and Science, 2009-j-19p, pp. 1-6.
- [6] O. Morikawa and T. Maesako, “HyperMirror: Toward Pleasant-to-Use Video Mediated Communication System,” Proc. ACM Conf. Computer-Supported Cooperative Work (CSCW98), ACM Press, 1998, pp. 149-158.
- [7] A. Kimura, T. Kuroda, Y. Manabe, and K. Chihara. “A study of display of visualization of motion instruction supporting,” Japan Journal of Educational Technology research, Vol.30, 2007, pp. 45-51.
- [8] MMDAgent: Toolkit for building voice interaction systems <http://www.mmdagent.jp/>, last access: Dec. 6, 2011.



Figure 3. Skelton information

TABLE I. GESTURE INDEX

Gesture Index		
Command	Effect	Pose
JMenu	display menu	Raise the right hand, specify by the left hand
JRotate	Rotate the camera view	Raise the left hand, specify by the right hand
JMove	Move the camera view	Raise the left hand, specify by the right hand
JPopItem	Create the tact	Put hands together

## Situated Cognitive Engineering: The Requirements and Design of Automatically Directed Scenario-based Training

Marieke Peeters, John-Jules Ch. Meyer  
*Intelligent Systems Group*  
*Utrecht University*  
*Utrecht, The Netherlands*  
 {M.M.M.Peeters, J.J.C.Meyer}@uu.nl

Karel van den Bosch  
*Training and Performance Innovations*  
*TNO*  
*Soesterberg, The Netherlands*  
 Karel.vandenBosch@tno.nl

Mark A. Neerincx  
*Man-Machine Interaction*  
*Delft University of Technology*  
*Delft, The Netherlands*  
 M.A.Neerincx@tudelft.nl

**Abstract**—Serious games enable trainees to practice independently of school, staff, and fellow students. This is important as amount of practice directly relates to training efficacy. It is also known that personalized guidance elevates the benefits of training. How to achieve automated guidance, for example to be used in serious games, is a yet unsolved issue. This paper uses the situated Cognitive Engineering method to analyze the operational demands, theoretical foundations and technological opportunities for the design of an automatically directed scenario-based training system (AD-SBT). AD-SBT guides training by selecting scenarios that match the trainee’s competency level, by monitoring the training process, and by offering appropriate support. Three instructional principles are used: adapt training to the trainee’s cognitive characteristics, strengthen the trainee’s will to learn, and foster transfer of learned skills. This paper reports evidence taken from the literature and by means of a use case simulation to validate and verify the presented requirements for AD-SBT and the underlying claims. Results show that the introduced requirements baseline and the resulting design for AD-SBT form a good starting point for future refinement and prototyping.

**Keywords**-director; scenario; training; cognitive engineering

### I. INTRODUCTION

Some professions (e.g., firemen, policemen, nurses) involve complex skills: integrated physical abilities, as well as cognitive abilities such as situation assessment and decision making. The most straightforward way to train complex skills is to have trainees engage in representative, real-life situations. However, research has shown that in order for training to be effective, trainees need guidance during skill acquisition [1]. Such guidance is usually offered by an instructor and in some cases even by an entire team of staff members, resulting in high costs of training in terms of money, time and resources. By simulating not only the learning environment, but also the process of guidance, these costs can be reduced.

To gain some understanding in the ways in which instructors normally guide their trainees, let us consider a clinical psychologist whose main tasks are to diagnose and treat patients suffering from mental health problems. These tasks are to be performed under varying circumstances: new

patients, patients halfway their treatment, all suffering from various disorders, all in need of different treatments, etc. Not every case is equally suitable for a trainee to learn from. How suitable a training case is for a particular trainee depends on the trainee’s experience and the complexity of the case.

Say, a trainee needs to practice his first intake interview in an on-the-job training setting. However, the first patient happens to be suicidal. In such a case, the instructor may decide to let this patient pass as a training case, because the trainee lacks the experience to treat this patient and to deal with the additional emotional responsibility. The next patient, a boy having trouble mourning over the death of his father, is probably a better case to start with. During training an instructor repeatedly selects an appropriate case based on the trainee’s latest performance. The ultimate goal is to guide the trainee to a performance level that enables him to diagnose and treat all possible cases.

If training does not take place on-the-job, but in a role-played scenario, the instructor is able to refine the case even more, by playing the patient himself or by hiring an actor to play the patient while the instructor functions as a director. Either way, the instructor is now able to manipulate the scenario as it develops, by controlling the patient’s behavior.

Selecting the appropriate learning task (scenario) and the appropriate amount of support (manipulative directions) are powerful ways to guide the trainee and create effective training situations. To be able to add automated guidance to virtual learning environments, such as serious games, it is necessary to uncover the principles underlying training task selection and manipulation. How do instructors decide what task is appropriate for a trainee to learn from? What events cause instructors to intervene in a training scenario?

This paper uses the situated Cognitive Engineering (sCE) method to achieve design principles and an architecture for automatically directed scenario-based training (AD-SBT) [2]. The sCE method has been previously applied to the defense and space domain for the design of systems involving computer-supported task performance. Its roots lie in the field of cognitive engineering, a science of user-centered design which aims to uncover the principles behind human

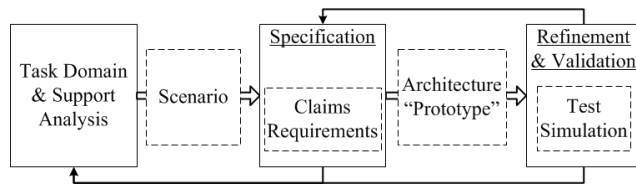


Figure 1: The situated Cognitive Engineering method [2]

action and performance that are relevant for the design of systems comprising of both people and machines, in order to maximize the performance of this joint system [3], [4].

The sCE method (Figure 1) consists of a process of iterative development cycles. Each cycle is directed at refining the system's specifications in more and more detail. The system's initial specifications follow from careful analysis of the task domain, human factors knowledge (e.g., relevant theories, guidelines and support concepts) and technological opportunities. This set of specifications, consisting of claims, requirements and use cases, forms the foundation for a prototype. This set is to be refined by evaluating the prototype. The entire development process is situated, meaning that it takes place within the context of the domain.

The current paper describes the first development cycle of the sCE method, resulting in a requirements baseline, and thereby it is structured as follows. Section II specifies the analysis of the task domain, the human factors knowledge and the technological opportunities for the intended AD-SBT system. Section II-D presents a scenario illustrating the intended use of the system. The drivers identified in Section II form the foundation for the system's specification in terms of claims and requirements presented in Section III. Section IV presents the proposed design architecture for a prototype of the system, which is based upon the aforementioned requirements. Section V describes possible ways (a test and a simulation) for (future) refinement of the requirements and verification of the architecture (prototype). Conclusions can be found in Section VI.

## II. TASK DOMAIN AND SUPPORT ANALYSIS

This section covers the results of our domain and support analysis, leading to the operational, theoretical and technological drivers for AD-SBT. The operational driver, a training method called scenario-based training, is presented. Subsequently, it is shown how SBT can be structured or *directed*, by using three principles of instructional design identified in the human factors literature. These principles form the theoretical driver. Lastly, the technological opportunities for automatically directed scenario-based training are discussed. These form the technological driver for AD-SBT. Finally, the envisioned system is illustrated by means of a scenario describing the intended use of the system.

### A. Operational demands: Scenario-based training (SBT)

Scenario-based training (SBT) has proven to be a powerful training method [5], [6]. It is consistent with the principles

recognized in dominant instructional theories [7]. During SBT, trainees prepare, execute and evaluate training scenarios, i.e., real-life, relevant and meaningful storylines within a simulated environment (SE) [5], [8]. The fidelity of the SE may vary, ranging from the actual task environment to highly symbolic representations thereof. Training within an SE has several benefits in comparison to on-the-job training: an SE can be prepared, controlled, reused, and improved upon, leading to the reduction of risks, costs, and resources.

Training scenarios generally address specific learning objectives. To ensure their realism and didactical value, scenarios are authored and prepared in advance of the training session by an instructor. Most often the scenario contains so-called non-player characters (NPCs) with whom the trainee needs to interact, such as teammates, officers, patients, or opponents. These roles may be played by the actual persons (in case of teammates for instance), by staff members and actors (especially in case of opponents or enemies), or by virtual actors such as intelligent agents.

### B. Theoretical foundation: Human factors knowledge

There are many instructional principles in the human factors literature. Three recurring and prominent principles relevant for AD-SBT are identified and described in the subsections below: 'Transfer - keeping an eye on the subject matter'; 'Adapt to the trainee's cognitive characteristics'; and 'Increase the trainee's will to learn'. These principles are used for a further specification of AD-SBT.

#### 1) Transfer - keeping an eye on the subject matter:

Effective training is supposed to lead to the display of the trained skills during future job performance (transfer) [9]. To promote transfer, training tasks should be *authentic*; i.e. represent the tasks that the trainee will perform in his future profession [10], [11]. Additionally, training tasks should come in a wide diversity; by generalizing solutions over various tasks, trainees learn to abstract away from them. This abstraction leads to the recognition of the underlying principles to be applied in the actual task environment [9].

#### 2) Adapt to the trainee's cognitive characteristics:

To facilitate learning, one needs to adapt the training to the trainee's competencies. Effective instruction takes limited working memory capacity into account. As the trainee's competencies grow during training, the amount of information that can be processed by the trainee's working memory slowly increases, allowing for the task load to increase [12]–[14]. This can be done by fading support (scaffolding); i.e. adjusting the amount of feedback, cues, simultaneous events or time restrictions [15]. Another way to increase the task load is by increasing task complexity; i.e. selecting tasks that require a little more than the trainee's current competencies.

A second aspect to keep in mind is that trainees make use of different learning strategies. This means that the instructor should be able to choose from different teaching styles, e.g., by means of examples, through learning by doing, by

applying knowledge to cases, by evaluating and reflecting on outcomes, by connecting experiences, etc. [16]

3) *Increase the trainee's will to learn*: The higher a trainee's motivation, the more efforts he will put into training and into transferring the trained competencies to the actual task environment [17]. Motivation is related to the level of *self-efficacy*: the trainee's truthful beliefs about his task performance capabilities [18]. Motivation can be intrinsic (engaging in an activity for its own sake) or extrinsic (engaging in an activity as a means to an end), the former being the more favorable one, as it is positively related to trainee achievements in contrast to extrinsic motivation which is negatively related to trainee achievements [19], [20]. Intrinsic motivation is promoted by offering the trainee meaningful and relevant learning experiences [21].

As argued earlier (Section II-B2), training tasks need to be compatible with the trainee's competencies. In addition, the trainee also needs to believe in his ability to master the task (self-efficacy) in order for him to be motivated to perform the task. A balance must be found between the offered challenge and the trainee's competencies to prevent him from getting bored or frustrated.

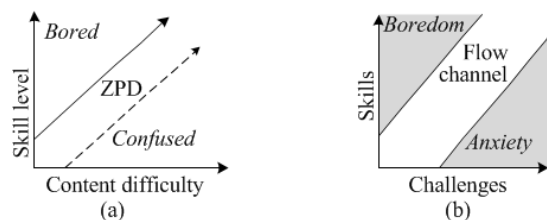


Figure 2: (a) Zone of Proximal Development (ZPD) [22] and (b) Flow [23]

Figure 2 displays two graphs representing the Zone of Proximal Development (ZPD) [22] and Flow [23]. Both figures refer to the zone between anxiety/confusion and boredom. Vygotsky's ZPD [24] is a common concept in instructional theory, representing a class of training tasks slightly more complex than the tasks the trainee is currently able to perform. Such tasks challenge the trainee to develop new skills or insights. The ZPD shows a remarkable resemblance to flow, a common concept in game research [23]. Flow is defined by Csikszentmihalyi (1991) as a mental state that results from appropriate challenge and truthful self-efficacy and is characterized by high levels of motivation, concentration and enjoyable learning. Recently, flow as a concept has been adopted by game-researchers [25], [26]. An interesting range of experiments has been published about the derivation of trainee's mental state - i.e. excitement, frustration, boredom, pleasure, challenge, interest, and even flow - by using psycho physiological measurements [27]–[29]. The results of these experiments can be used to automatically adjust the amount of challenge to keep the trainee in flow.

### C. Technological opportunities

By implementing SBT into a virtual environment instead of a real one and using intelligent agent technology, it is possible to automate the NPCs [30] in such a way that they behave in a consistent and explainable fashion. This is preferable in a training system, since it results in believable and understandable NPCs. Several authors in the field of instructional design emphasize the potential of serious games in education [31]–[33], because they lead to an increase in motivation [20], [31], and because necessary adjustments to the environment and the behavior of the NPCs can be executed behind the scenes while the scenario unfolds [15]. In addition to the NPCs, we propose to introduce a director agent (DA). A DA creates suitable learning situations for the trainee by manipulating the NPC-behaviors behind the scenes in real-time.

There have been other proposals for DA architectures before, particularly within the domain of interactive narrative. One example is the Interactive Storytelling Architecture for Training (ISAT) of Magerko et al. (2005) [34], which uses partial order planning to generate variable storylines. Another example is the framework by Si, Marcella and Pynadath (2009). They describe a DA using their framework 'Thespian', which includes a runnable user model that can predict the trainee's future behavior [35]. Moreover, the agents in Thespian are able to reason about other agents' intentions. This DA intervenes by changing NPCs' beliefs and plans, possibly resulting in inconsistent NPC behaviors, which can be harmful for training since trainees may not accept these characters as realistic or believable. IN-TALE is a third example of such a framework. It was proposed by Riedl et al. (2008) [36] and should lead to believable failure of the agents' plans in case of conflicts between the NPC's plans and the DA's directions. The mentioned architectures generally focus on drama management. Although they acknowledge that their architectures also offer opportunities for training, the actual implementation of instructional theory is mostly neglected.

### D. Illustration of the envisioned AD-SBT system

To illustrate the functionalities of the envisioned system, a description of a scenario in the domain of clinical psychology is given below.

Karen and Luke, two psychology students, start the training session. Karen plays the patient (NPC) while Luke plays the psychologist. The DA selects a scenario that fits Luke's learning goals, i.e. 'thorough questioning' and 'conversational management'. Karen plays a woman with bulimic disorder in denial of her problem, who was sent by her doctor. Karen follows the DA's instructions: 'change your position constantly', 'talk about your indignation for being sent to a therapist', and 'do not talk about food nor your figure'.

The DA receives updates about the conversation. Karen and Luke have been talking about Karen's household chores for a while. Luke



has not asked any questions regarding the reasons that made Karen's doctor decide to redirect her to a psychologist. Moreover, Luke is not showing the right body language. Based on this information, the DA adjusts its instructions to Karen and its feedback to Luke. Karen is instructed to calm down and tell Luke why she visited her doctor. Luke receives feedback on his posture and is reminded of his learning goals to offer him a better focus. After a while, the DA stops the scenario. It encourages Karen and Luke to discuss the case, and provides them with feedback and a concise overview of the training session.

### III. REQUIREMENTS OF AN AD-SBT SYSTEM

In Section II the design space for AD-SBT was presented. The current section presents the requirements for the AD-SBT system, based on this design space. During AD-SBT, training scenarios need to be automatically adapted to the trainee's cognitive and emotional state in real-time by a director agent (DA). The DA must create a wide variety of authentic training scenarios that match the trainee's needs.

A simplified list of requirements (R1 – R5) for AD-SBT is given along with the claims (C1.1 – C5.2) that should justify each of them, as they have become available by applying the sCE method. The requirements are justified when the underlying claims are validated by means of existing empirical evidence, simulations and future research.

- R1 Match scenarios to the trainee's current competencies.
  - C1.1 Presenting scenarios in order of increasing complexity and matching them to the trainee's level of experience prevents cognitive overload (see Section II-B2).
- R2 Adjust the level of scaffolding based upon the trainee's emotional responses.
  - C2.1 Adjusting the level of challenge fosters flow and high levels of motivation (see Section II-B3).
- R3 Create realistic (authentic) scenarios.
  - C3.1 Authentic training tasks foster transfer (see Section II-B1).
  - C3.2 Authentic training tasks foster motivation (see Section II-B3).
  - C3.3 Engaging in authentic training tasks fosters immersion, and thereby flow and motivation.
- R4 Create variable scenarios.
  - C4.1 This will foster transfer (see Section II-B1).
- R5 Provide feedback about interventions in the scenario.
  - C5.1 This will foster self-efficacy (see Section II-B3).
  - C5.2 This will foster a better understanding of the task domain.

### IV. THE DESIGN OF AN AD-SBT SYSTEM

In the current section the design architecture for AD-SBT (Figure 3) is presented. This architecture is based on the requirements listed in Section III. The architecture consists of two processes. Both of these adapt the scenario to fit it to the trainee's needs by instructing the agents to change their behaviors.

The first adaptation process is a *reactive process* - depicted by the solid line - that reacts upon psycho physiological measurements (bottom right) by adjusting the level of scaffolding to prevent the trainee from getting bored or frustrated. A reasoning engine (top right star) decides whether an adjustment in of the scaffolding (support/challenge) level is necessary. If so, it determines the nature of the adjustment to be made and sends this to the user model and to the agents

controlling the characters and environmental elements, so they are able to change their behaviors accordingly.

The second process is a *reasoning process* (grey interrupted lines), brought about by the DA (top left circle). To decide on how to continue the scenario, the DA (1) uses a didactic reasoning engine (top left star), (2) an expert model and a scenario model (the two hexagons), and (3) consults a user model (grey striped square) and a world model (white checkered square). The user model contains the trainee's achievements, performed tasks, and reached checkpoints. The DA uses its didactic reasoning engine to reason about the user model and the expert model and decides what learning goals are suitable for the trainee. When an adjustment of the learning goals is appropriate, the DA sends a notification to the feedback engine, which should generate proper explanations for the adjustments to communicate to the trainee through the environment agent (Agent3 in Figure 3).

Once the DA has selected the learning goals, it uses the scenario database to select scenes that address these learning goals, after which it consults the world model to see what scene can be implemented without interrupting the believability of the storyline. The world model is updated with information coming from the agents and the user model (the grey dotted lines) and keeps track of all relevant facts in the world. The DA consults the world model to see what scenes fit the current world state. The DA randomly selects a scene from the resulting set - consisting of scenes appropriate for the trainee and matching the world state - and sends it to the other agents, so they are able to change their behaviors accordingly.

Note that this division into two adjustment cycles is not entirely new. VanLehn (2006) also describes two loops in the behavior of intelligent tutoring systems (ITSs); an inner and an outer loop [14]. He suggests that the outer loop is used to determine what learning task the ITS is supposed to offer next, whereas the inner loop is used to select the steps within the task, taking feedback, support and assessment into account.

### V. (FUTURE) REFINEMENT AND VALIDATION

There are several ways to validate the claims presented in Section III. Two examples of claim validations will be presented as a further clarification of the sCE method: a test (described in [15]) and an outline for a possible use case simulation.

#### A. Test

An explorative study into the applicability of this architecture was conducted. For this study an AD-SBT system applied to the domain of 'in-company emergency services' was created. 'In-company emergency services' refer to a team of employees within a company who are trained to provide first-aid, extinguish small fires or clear the office

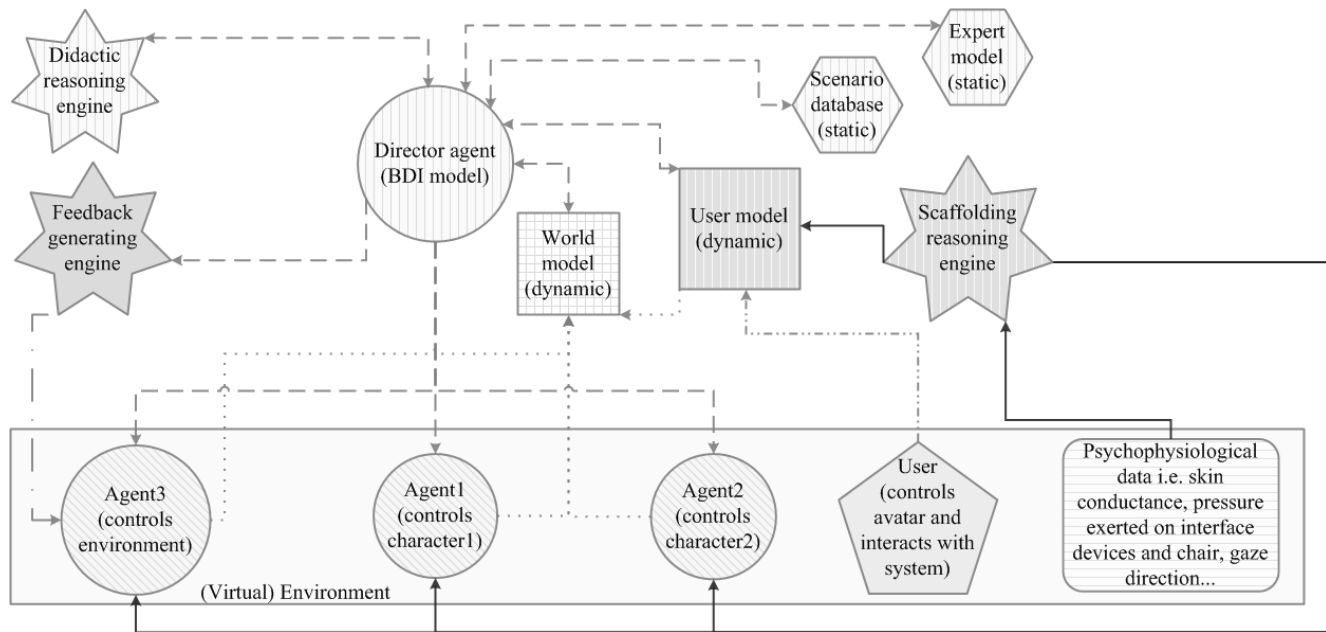


Figure 3: The Director Agent Architecture

building in case of emergencies, until the official emergency services have arrived.

Because of the cumbersome task of implementing a prototype of the envisioned training system, a Wizard of Oz prototype was developed. The simulated environment was not virtual, the NPCs were human actors and the director was also human. This created the possibility to investigate approaches for directing training without the need to actually create a virtual environment with added intelligent agent technology. Two actors received training to play the role of the NPCs in four scenarios. The human director was able to adjust the level of scaffolding by intervening in the behavior of the NPCs using a very detailed script and in-ear portophones. The aim of the interventions was to adjust the level of scaffolding to match the performance of the trainee. All of the resulting scenarios were recorded on video.

Subsequently, an experiment was conducted to test for the effects of the director's interventions upon the learning value of the scenario. Professional instructors from the domain judged twenty video fragments coming from the recordings of the study described above. In ten of these fragments, the scenarios were directed, meaning that the level of scaffolding could be adjusted in real-time according to a script. The other ten fragments came from undirected scenarios, meaning that the scenarios were not adjustable in real-time. The instructors were under the impression that all fragments were undirected and individually judged the learning value at set points in time.

The results of this experiment indicated that the DA's interventions resulted in more suitable learning situations compared to the undirected scenarios. These results support the validation of claim C2.1 and the justification of

requirement R2 (Section III). Moreover, the trainees reported AD-SBT to be realistic, motivating and stimulating during interviews, thereby supporting not only the validation of claim C2.1 and justifying requirement R2, but also claim C3.2 and requirement R3. For a more extensive description of this experiment, the reader is directed to [15].

### B. Use case simulation

It is emphasized that the following use case simulation is fictional. It serves as an example of how a design architecture could be verified by means of simulation through use cases. Formal use cases, however, need a far more detailed specification and should be constructed in consultation with domain experts.

#### Use case simulation 1

- 1) A trainee is currently enrolled in scene S14. He is receiving full support. His learning goal is G9.
- 2) The psycho physiological data imply stress indications, meaning that the trainee is in need of more support.
- 3) The scaffolding engine sends an update to the user model: Unable to adjust the level of scaffolding in the desired direction.
- 4) The director agent selects a set of less complicated, suitable learning goals {G2, G3 and G6} by reasoning about the expert and user model.
- 5) The world model contains conditions {C1, C3 and C14}.
- 6) Scene S9 and scene S11 in the scenario database both involve suitable learning goals, as selected in step 4.
- 7) The director agent sends a notification about the adjustment of the current learning goals to the feedback generating engine.
- 8) Both of the scenes proposed in step 6 are applicable, since their preconditions match the conditions in the world model.
- 9) The director agent randomly selects scene S11 from the set of applicable scenes as computed in step 6 and step 8.
- 10) The director agent notifies the agents about the scene selection.
- 11) The feedback generating engine sends an explanation for the adjustment of the learning goals to the environment agent.

- 12) The agents adjust their behavior to match scene S11.
- 13) The environment agent explains the new learning goals to the trainee.

This use case simulation exemplifies how the DA adjusts the learning task to address the appropriate learning goals, thereby verifying requirement R1 in Section III. The DA uses the world model to check whether the scenes are compatible with the current world state, thereby ensuring a realistic course of events, which verifies R3. Moreover, the DA creates variable scenarios by randomly selecting a scene from a set of options, thereby verifying R4. Finally, the DA provides feedback about the chosen learning path, which verifies R5.

## VI. CONCLUSION AND FUTURE WORK

Forms of training that allow students to practice independently (such as in serious games) are extremely valuable as they allow for higher frequencies and volumes of training. In this paper we argue that the benefits can be increased if such environments are equipped with a system that is able to provide the trainee with guidance: ‘automatically directed scenario-based training’ (AD-SBT). AD-SBT can be described as scenario-based training extended with a director agent. A director agent is able to manipulate the scenario in real-time to realize personalized interventions, guidance and support during scenario-based training.

The situated Cognitive Engineering method is used to present a design rationale for AD-SBT. This design rationale consists of (1) the operational demands, (2) relevant principles coming from instructional theory and game research and (3) the envisioned technology. The use of the method of situated Cognitive Engineering led to the identification of requirements and associated claims for an AD-SBT system. These requirements and claims (Section III) are based on knowledge about transfer, motivation, cognitive overload, and metacognitive skills. The requirements form the foundation for the ‘automatically directed scenario-based training’ (AD-SBT) architecture presented in Section IV. The architecture specifies two main processes: a reactive process controlling the level of scaffolding and a reasoning process controlling the complexity of the scenario.

A first test, described to a greater detail in [15], showed that AD-SBT significantly improves the learning value of the scenario, by being able to adjust the scenario in real-time to match the trainee’s current needs. Moreover, trainees reported high levels of motivation during the AD-SBT scenarios. These results suggest that the requirements outlined in Section III forms a good baseline for further refinement in future research and prototyping.

Earlier research on intelligent tutoring systems has mainly focused on structuring and ordering the learning material to support the construction of well embedded and coherent knowledge structures. This guidance and support helps the trainee to comprehend the lessons to be learned. However,

research on intelligent tutoring systems has neglected the importance of storylines and contextualization of learning materials to add meaning to the learning materials, to relate to the trainee’s expectations and motivation, and to increase the trainee’s will to learn. In contrast, research on serious games has mainly focused on engagement, motivating the trainee to learn, fostering transfer, and creating realistic learning environments. But serious games lack the structured approach to learning that is so important to guide the trainee during training. The research described here is a promising step in the development of a training system that does not only engage the trainee in active learning through participation in a storyline, but also provides the necessary guidance and support to the trainee. This research therefore helps to link various fields within the training research domain, such as serious games, scenario-based training, interactive storytelling, intelligent tutoring systems and instructional theory.

The next development cycle in the situated Cognitive Engineering process is directed at a partial implementation of the AD-SBT system. As the main focus of the research described here is automatic guidance during scenario-based training, this means that the reasoning determining the interventions will be modeled and implemented first. In order to model the reasoning process, we have to deal with a number of issues, such as: ‘Is it possible to translate didactic strategies to generic rules in a didactic reasoning engine?’, ‘About what sorts of trainee behaviors should the director agent be able to reason?’, ‘What types of interventions does a human instructor perform and how can they be automated?’, ‘What knowledge about the trainee should be taken into account during the reasoning process?’, and ‘How much knowledge is needed about the possible behaviors of the non-player characters to be able to reason about and to execute effective interventions?’. The modeling, implementation, and (experimental) evaluation of the reasoning process will bring us another step closer to a fully functioning AD-SBT system.

## REFERENCES

- [1] P. Kirschner, J. Sweller, and R. Clark, “Why minimal guidance during instruction does not work,” *Educ psychol*, pp. 75–86, 2006.
- [2] M. Neerinx and J. Lindenberg, “Situated cognitive engineering for complex task environments,” in *Naturalistic Decision Making and Macrocognition*. Ashgate, 2008, pp. 373–390.
- [3] D. Norman, “Cognitive engineering,” in *User centered system design*. Lawrence Erlbaum Associates, 1986, pp. 31–61.
- [4] J. Rasmussen, *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. Elsevier Science Inc., 1986.
- [5] R. Oser, “A structured approach for scenario-based training,” in *HFES Ann. Meeting*, 1999, pp. 1138–1142.

- [6] E. Salas, H. A. Priest, K. A. Wilson, and A. B. Adler, "Scenario-based training: Improving military mission performance and adaptability," in *Military Life*. Praeger Publishers, 2006, pp. 32–53.
- [7] M. D. Merrill, "First principles of instruction," *ETR&D*, pp. 43–59, 2002.
- [8] J. Cannon-Bowers, J. Burns, E. Salas, and J. Pruitt, "Advanced technology in scenario-based training," in *Making Decisions Under Stress*. APA, 1998, pp. 365–374.
- [9] T. Baldwin and J. Ford, "Transfer of training: A review and directions for future research," in *The training and development sourcebook*. HRD Pr, 1994.
- [10] R. Grabinger and J. Dunlap, "Rich environments for active learning: a definition," *ALT-J*, pp. 5–34, 1995.
- [11] M. Young, "Instructional design for situated learning," *ETR&D*, pp. 43–58, 1993.
- [12] J. Van Merriënboer, P. Kirschner, and L. Kester, "Taking the load off a learner's mind," *Educ Psychol*, pp. 5–13, 2003.
- [13] J. van Merriënboer and J. Sweller, "Cognitive load theory and complex learning," *Educ Psychol Rev*, pp. 147–177, 2005.
- [14] K. VanLehn, "The behavior of tutoring systems," *IJAIED*, pp. 227–265, 2006.
- [15] M. Peeters, K. v. d. Bosch, J.-J. Meyer, and M. Neerincx, "Scenario-based training: director's cut," in *AIED conference*, 2011, pp. 264–272.
- [16] D. Orlich, R. Harder, R. Callahan, M. Trevisan, and A. Brown, *Teaching strategies: A guide to effective instruction*. Wadsworth Pub Co, 2009.
- [17] D. Schunk, P. Pintrich, and J. Meece, *Motivation in education*. Pearson/Merrill Prentice Hall, 2009.
- [18] E. Linnenbrink and P. Pintrich, "Motivation as an enabler for academic success," *School Psychol Rev*, pp. 313–327, 2002.
- [19] M. Lepper, J. Corpus, and S. Iyengar, "Intrinsic and extrinsic motivational orientations in the classroom," *J Educ Psychol*, pp. 184–196, 2005.
- [20] T. Malone, "Toward a theory of intrinsically motivating instruction," *Cog Sci*, pp. 333–369, 1981.
- [21] M. Mitchell, "Situational interest: Its multifaceted structure in the secondary school mathematics classroom," *J Educ Psychol*, pp. 424–436, 1993.
- [22] T. Murray and I. Arroyo, "Toward measuring and maintaining the zone of proximal development in adaptive instructional systems," in *ITS conference*, 2002, pp. 749–758.
- [23] M. Csikszentmihalyi, *Flow: The psychology of optimal experience*. Harper Collins, 1991.
- [24] L. Vygotsky, *Mind in society*. Harvard University Press, 1978.
- [25] W. Admiraal, J. Huizenga, S. Akkerman, and G. Dam, "The concept of flow in collaborative game-based learning," *Comput Hum Behav*, pp. 1185–1194, 2011.
- [26] P. Sweetser and P. Wyeth, "Gameflow," *Computers in Entertainment*, 2005.
- [27] A. Kapoor, W. Burleson, and R. W. Picard, "Automatic prediction of frustration," *Int J Hum-Comput St*, pp. 724–736, 2007.
- [28] S. Mota and R. W. Picard, "Automated posture analysis for detecting learner's interest level," in *CVPRW conference*, 2003, p. 49.
- [29] L. Nacke and C. A. Lindley, "Flow and immersion in first-person shooters," in *Future Play Conference*, 2008, pp. 81–88.
- [30] K. van den Bosch, M. Harbers, A. Heuvelink, and W. van Doesburg, "Intelligent agents for training on-board fire fighting," in *DHM conference*, 2009, pp. 463–472.
- [31] M. Dickey, "Engaging by design," *ETR&D*, pp. 67–83, 2005.
- [32] S. Egenfeldt-Nielsen, "Overview of research on the educational use of video games," *Digital Kompetanse*, pp. 184–213, 2006.
- [33] L. Rieber, "Seriously considering play," *ETR&D*, pp. 43–58, 1996.
- [34] B. Magerko, R. E. Wray, L. S. Holt, and B. Stensrud, "Customizing interactive training through individualized content and increased engagement," in *IITSEC conference*, 2005.
- [35] M. Si, S. Marsella, and D. Pynadath, "Directorial control in a decision-theoretic framework for interactive narrative," in *Interactive Storytelling conference*, 2009, pp. 221–233.
- [36] M. Riedl, A. Stern, D. Dini, and J. Alderman, "Dynamic experience management in virtual worlds for entertainment, education, and training," *ITSSA, Special Issue on Agent Based Systems for Human Learning*, pp. 23–42, 2008.

## Augmented Reality in Minimally Invasive Surgery

Lucio Tommaso De Paolis  
 Department of Innovation Engineering  
 University of Salento  
 Lecce, Italy  
 lucio.depaolis@unisalento.it

Giovanni Aloisio  
 Department of Innovation Engineering  
 University of Salento  
 Lecce, Italy  
 giovanni.aloisio@unisalento.it

**Abstract**—The advantages of the Minimally Invasive Surgery are evident for the patients, but these techniques have some limitations for the surgeons. In medicine, the Augmented Reality (AR) technology allows surgeons to have a sort of “X-ray” vision of the patient’s body and can help them during the surgical procedures. In this paper we present two applications of Augmented Reality that could be used as support for a more accurate preoperative surgical planning and also for an image-guided surgery. The first AR application can support the surgeon during the needle insertion for the Radiofrequency Ablation of the liver tumours in order to guide the needle and to have a precise placement of the instrument within the lesion. The augmented visualization can avoid as much as possible to destroy healthy cells of the liver. The second AR application can support the surgeon in the preoperative surgical planning by means of the visualization of the 3D models of the organs built from patient’s medical images and in the choice of the best insertion points of the trocars in the patient’s body.

**Keywords** - Augmented Reality, medical images, minimally invasive surgery, laparoscopic pediatric surgery, RF ablation

### I. INTRODUCTION

One trend in surgery is the transition from open procedures to minimally invasive interventions, where visual feedback to the surgeon is only possible through the laparoscope camera and direct palpation of organs is not possible. Minimally Invasive Surgery (MIS), such as laparoscopy or endoscopy, has changed the way to practice the surgery and, as a promising technique, the use of this surgical approach is nowadays widely accepted and adopted as an alternative to classical procedures.

These techniques offer the possibility to surgeons of reaching the patient’s internal anatomy in a less invasive way and causing only a minimal trauma. The diseased area is reached by means of small incisions in the body and specific instruments and a camera are inserted through these ports; during the operation a monitor shows what is going on inside the body.

Shorter hospitalizations, faster bowel function return, fewer wound-related complications and a more rapid return to normal activities have contributed to accept these surgical procedures and, if the advantages of this surgical method are evident on the patients, these techniques involve some limitations for the surgeons. In particular, the imagery is in

2D and the surgeon needs to develop new skills and dexterity in order to estimate the distance from the anatomical structures and work in a very limited workspace.

The acquisition of medical images (CT or MRI) associated to the latest medical image processing provides an accurate knowledge of the patient’s anatomy and pathologies and could lead to an improvement in patient care by guiding of the surgeons during the surgical procedure.

The emerging Augmented Reality (AR) technology has the potential to bring the direct visualization advantages of open surgery back to minimally invasive surgery and can increase the physician’s view with information gathered from the patient’s medical images.

AR technology refers to a perception of a physical real environment whose elements are merged with virtual computer-generated objects in order to create a mixed reality. In addition to a mixture of real and virtual information, an AR application has to run in real time and its virtual objects have to be aligned (registered) with real world structures. Both of these requirements guarantee that the dynamics of real world environments remain after virtual data has been added [1].

In medicine the AR technology makes it possible to overlay virtual medical images of the organs on the real patient and it allows the surgeon to have a sort of “X-ray vision” of the patient’s internal anatomy. In order to register the data and fuse virtual and real imagery in real time, special devices are used in the AR platform.

An advantage of using AR in surgery could be found in a better spatial perception and in the duration of the surgical procedure that could be shorter than in conventional way.

The aim of this paper is to present two AR applications that could be used as support for a more accurate surgical preoperative planning and also for an image-guided surgery. The first application can support the surgeon during the needle insertion in the radiofrequency ablation of the liver tumours; the second one allows the surgeon to choose the more appropriate points for the insertion of the trocars a pediatric laparoscopic procedure.

### II. PREVIOUS WORKS

Motivated by the benefits that MIS can bring to patients, many research groups are now focusing on the

development of systems in order to assist the surgeons during the minimally invasive surgical procedures.

Furtado and Gersak [1] present some examples of how AR can be used to overcome the difficulties inherent to MIS in the cardiac surgery.

Samset et al. [3] present some decision support tools based on concepts in visualization, robotics and haptics and provide tailored solutions for a range of clinical applications.

Bichlmeier et al. [4] focus on the problem of misleading perception of depth and spatial layout in medical AR and present a new method for medical in-situ visualization.

Navab et al. [5], [6] present a new solution for using 3D virtual data in many AR medical applications and introduce the concept of a laparoscopic virtual mirror, a virtual reflection plane within the live laparoscopic video.

De Paolis et al. [7] present an AR system that can guide the surgeon in the operating phase in order to prevent erroneous disruption of some organs during surgical procedures. They provide distance information between the surgical tool and the organs and they use a sliding window in order to obtain a more realistic impression that the virtual organs are inside the patient's body.

Nicolau et al. [9] present a real-time superimposition of virtual models over the patient's abdomen in order to have a three dimensional view of the internal anatomy. The authors have used the developed system in an operating room and to reduce the influence of the liver breathing motion they have tried to simulate the patient's breathing cycle.

LiverPlanner [10], [11] is a virtual liver surgery planning system developed at Graz University of Technology that combines image analysis and computer graphics in order to simplify the clinical process of planning liver tumor resections. The treatment planning stage enables the surgeon to elaborate a detailed strategy for the surgical intervention and the outcome of pre-operative planning can then be used directly for the surgical intervention.

Maier-Hein et al. [12] present a system developed for computer-assisted needle placement that uses a set of fiducial needles to compensate for organ motion in real time; the purpose of this study was to assess the accuracy of the system in vivo.

Stüdeli et al. [13] present a system that provides surgeon, during placement and insertion of RFA needle, with information from pre-operative CT images and real-time tracking data.

### III. THE USED TECHNOLOGIES

A reconstruction of the 3D model of the anatomical and pathological structures of the patient is required in order to improve the standard slice view. An efficient 3D reconstruction of the patient's organs is generated by applying some segmentation and classification algorithms to medical images (CT or MRI) of the patient. The grey levels in the medical images are replaced by colours that are associated to the different organs [14].

Nowadays there are different software used in medicine for the visualization and the analysis of scientific images

and the 3D modelling of human organs; Mimics [15], 3D Slicer [16], ParaView [17], and OsiriX [18] play an important role among these tools.

In our application we have used 3D Slicer for the building of the 3D model of the patient's organ; 3D Slicer is a multi-platform open-source software package for visualization and image analysis and in Fig. 1 is shown a 3D model built from CT images of a two-year-old child with a benign tumour of the right kidney.

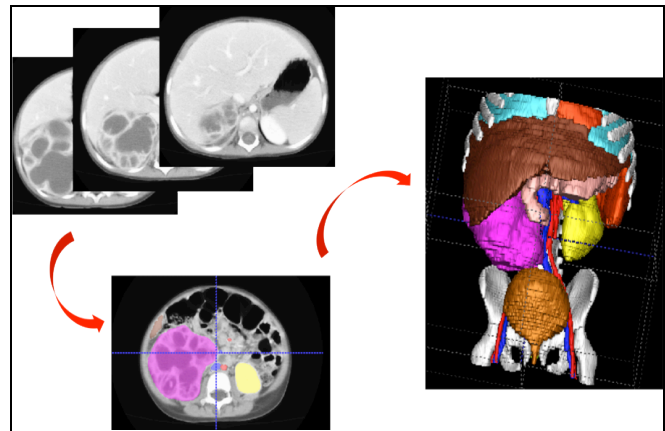


Figure 1. 3D model from the CT dataset

In the developed AR platform it is also necessary to use an optical tracker that is able to detect some intentionally introduced retro-reflective spheres (fiducials) placed on the surgical tools. In this way it is possible to know the right position and the orientation of the surgical instruments used during the surgical procedure. The tracking system is also used in order to permit the overlapping of the virtual organs on the real ones providing an augmented visualization of the scene. This registration phase is one of the most delicate step in an AR system for surgery.

We have used the Polaris Vicra optical tracker that provides precise and real-time spatial measurements of the location and orientation of an object or tool within a defined coordinate system [19].

For the visualization the Image-Guided Software Toolkit (IGSTK) that is a set of high-level components integrated with low-level open source software libraries and application programming interfaces has been used. IGSTK builds on the InsightToolkit (ITK) and the Visualization Toolkit (VTK) [20].

### IV. AR FOR THE RF ABLATION OF THE LIVER TUMOUR

#### A. The Liver Tumour

Hepatic cancer is one of the most common solid cancers in the world. Hepatocellular carcinoma (HCC) is the most common primary hepatic cancer. Unless primary hepatic cancer is quite rare, liver is second only to lymph nodes as a common site of metastasis from other solid cancers [21]. The liver is often the site of metastatic disease, particularly in patients with colorectal adenocarcinoma.

The use of chemotherapy for malignant form of liver cancer rarely led to good results in long-term survival rate. We have also to consider that chemotherapy produces negative effects in the lifestyle of the patient.

Today surgery is the best approach to avoid the death of the patient and the reversion of hepatic cancer. Unfortunately only from 5 to 15 per cent of HCC or hepatic metastasis patients diagnosed again undergo a potentially curative resection of the liver cancer [22], [23].

Patients with confined disease of the liver could not be candidates to resection because of multifocal disease, proximity of tumor to key vascular or biliary structures that preclude a margin-negative resection potentially unfavourable in case of presence of multiple liver metastases. Very often the tumor is also associated to a pre-existent cirrhosis that can reduce resection margins.

Liver transplant is the only radical therapy that eliminates the risk of recurrence but it can't be always used. So, since most of patients with primary or malignancies confined metastatic at the liver are not candidates for surgical resection, we should use different approaches to control and potentially cure liver diseases.

The Liver Radiofrequency Ablation (RFA) is a technique used since 1980's and consists in the placement of a needle inside the liver parenchyma in order to reach the centre of the tumor lesion. When the lesion is reached, an array of electrodes is extracted from the tip of the needle and it is that expanded in the tumor tissue. From these electrodes is injected for a certain time a radiofrequency current in the tumor tissue that causes tumor cell necrosis for hyperthermia (the local temperature is higher than 60 °C and cancer cells are more sensitive to heat than normal cells).

One problem in using radiofrequency tumor ablation technique is the correct placement of the needle that should reach the tumor lesion. Today surgeons use ultrasound, CT or MNR acquired during the needle placement in order to correctly direct the needle to the tumor. The use of these two-dimensional images makes the procedure very difficult and requires sometimes more than one insertion.

There is a need of enhancing the prognosis by a better control of the area interested by the RFA and the accuracy of the needle insertion in order to avoid some areas (important vessels) during the intervention.

A sort of guidance of the needle in tumour ablation procedures can be obtained using Augmented Reality technology. With the superimposition of the virtual models of the patient's anatomy (liver, cancer, etc) exactly where are the real ones, it is possible to make the needle placement less difficult.

### B. AR in RF Ablation of the Liver Tumours

The purpose of this AR application is to provide a guidance system that can help the surgeon during the needle insertion in liver RFA.

To overlap the virtual tool over the real one we track the position and the orientation of the real tool using some reflective spheres detected by means of an optical tracker. In this way the position and the orientation of the real tool are

measured and used in IGSTK to set the position and the orientation of the virtual tool.

To achieve a correct augmentation it is necessary to have a correspondence between the virtual organs and the real ones. This task is very difficult in an image-guided surgery application because a very small error in the registration phase can cause a serious consequence on the patient.

The registration process is carried out just before starting the surgical procedure and the applied method is based on the placement of fiducial points on the patient's body before the CT scanning and the detection of these in the 3D model built from the acquired medical images.

The optical tracker by means of a tool placed on the patient's body detects some possible movements of the patient over the operating table.

The application is provided of an user interface designed to be simple and functional at the same time. In the left side of that interface we have put the application control buttons; in the right-top window we show the 3D model and the augmented reality scene and in the right-bottom the three smaller windows where are placed the axial, coronal and sagittal views of CT dataset.

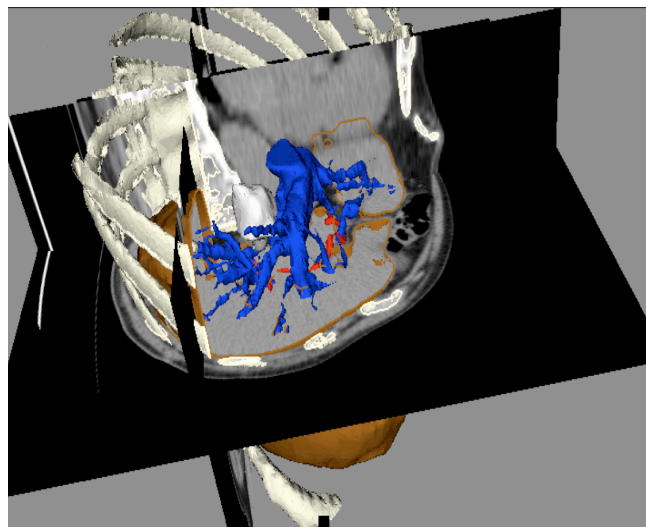


Figure 2. Clipping visualization applied to the liver and thoracic cage

There is the possibility to apply the clipping modality that permits the surgeon to dissect the model and to study its internal structure changing the opacity of the organs (Figure 2). The dissection could be made along the three principal axes.

The application features described till now are part of what we consider the pre-operative planning task. During this task the surgeon can use the application to study the pathology in a more simple and natural way than that provided by simple CT slice visualization.

For the navigation and augmentation task the surgeon needs to use the optical tracker and to carry out the registration task. When the registration process is complete, the virtual tool is shown in 3D view and is coupled with the

real one so it follows the movements that real tool makes.

In Figure 3 is shown the augmented visualization of the 3D virtual model over the patient's body (a dummy). This visualization can guide the surgeon during the needle insertion in the radiofrequency ablation of the liver tumour.

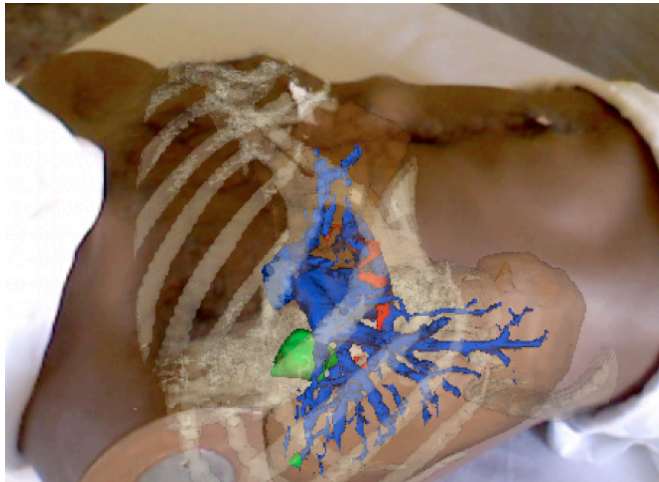


Figure 3. Augmented visualization in the RF ablation of the liver tumour

The application has been tested previously on a dummy and afterwards a first test has been carried out in the operating room in order to evaluate the system accuracy.

We are taking into account the possibility to include in the system also the simulation of the virtual model deformations due to the breathing of the patient.

#### V. AR FOR PEDIATRIC LAPAROSCOPY SURGERY

This application has the aim to provide support to the surgeons in a pediatric surgical procedure and, in particular, in the choice of the trocar entry points and the simulation of the interaction of the surgical instruments inside the patient's body.

Sometimes, using the standard insertion points for the surgical tools, also a simple surgical procedure can be very difficult due the difficulty to reach the organs. In this case the surgeon has to choose another insertion point in order to be able to carry out the surgical procedure in the most suitable way. Our aim is to avoid the occurrence of this situation and to provide a platform that allows to plan the insertion of the surgical tools on the virtual model of the patient's organs.

In this application we use the AR technology in order to visualize on the patient's body the precise location of the insertion points selected on the virtual model of the patient. Using the augmented visualization, the chosen entry points of the trocars are visualized on the patient's body in order to support the physician in the real trocar insertion phase.

For the augmented visualization, in order to have a correct and accurate overlapping of the virtual organs on the real ones, a registration phase is carried out; this phase is based on some fiducial points.

Figure 4 shows the specific section of the user interface

for the choice of the trocar insertion points.

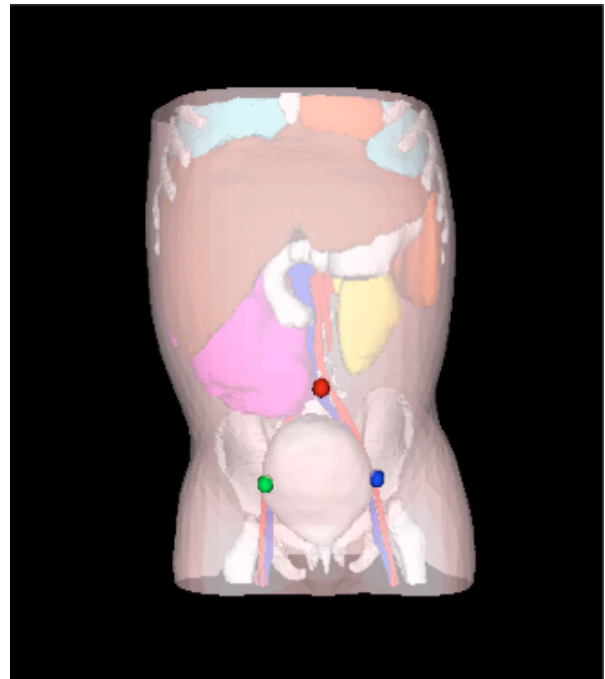


Figure 4. Interface for the choice of the trocar insertion points

Figure 5 shows the augmented visualization of the trocar's insertion points on the patient's body (a dummy).

The platform has been tested on two study cases (a two-year-old child with a benign tumor of the right kidney and a twelve-year-old child with a tumor of the peripheral nervous system) already operated by the surgeon; the future work will be the validation of the developed application on new real study cases.



Figure 5. Augmented visualization of the trocar's insertion points



VI. CONCLUSIONS AND FUTURE WORK

In this paper we present a guidance system for needle placement in radiofrequency ablation of the liver tumour and an AR platform for laparoscopic pediatric surgery.

The application for the RFA of the liver tumours has been tested in the laboratory on a specific testbed and, in order to start the validation of the system, a first test has been carried out in the operating room during an open surgery procedure for the liver tumour resection.

Operating rooms are overloaded of systems and devices and an efficient use of space is mandatory. In this first test the correct position of the devices close to the operating table has been chosen and a precise definition of the more appropriate fiducial points used for the registration phase has been decided. In Figure 6 is shown a phase of the first test in operating room.

As future work we are planning to add some improvements in the visualization of the organs taking into account the latest suggestions of the surgeons and to measure the precision of the image-guided application on a pig liver.



Figure 6. First test in operating room of the RF ablation application

REFERENCES

[1] S. Maad, "Augmented Reality. The Horizon of Virtual And Augmented Reality: The Reality of the Global Digital Age", Intech, January 2010, ISBN 978-953-7619-69-5.  
 [2] H. Furtado and B. Gersak, "Minimally Invasive Surgery and Augmented Reality. New Technology Frontiers in Minimally Invasive Therapies", 2007, pp. 195-201.  
 [3] E. Samset, D. Schmalstieg, J. Vander Sloten, A. Freudenthal, J. Declerck, S. Casciaro, Ø. Rideng, and B. Gersak, "Augmented Reality in Surgical Procedures", SPIE Human Vision and Electronic Imaging XIII, 2008.  
 [4] C. Bichlmeier, F. Wimmer, H. S. Michael, and N. Nassir, "Contextual Anatomic Mimesis: Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality", Proc. Sixth IEEE and ACM Int. Symposium on Mixed and

Augmented Reality (ISMAR '07), Nara, Japan, 2007, pp. 129-138.  
 [5] N. Navab, M. Feuerstein, and C. Bichlmeier, "Laparoscopic Virtual Mirror - New Interaction Paradigm for Monitor Based Augmented Reality", Proc. IEEE Virtual Reality Conf. 2007 (VR 2007), Charlotte, North Carolina, USA, 2007, pp. 10-14.  
 [6] C. Bichlmeier, S. M. Heining, M. Rustae, and N. Navab, "Laparoscopic Virtual Mirror for Understanding Vessel Structure: Evaluation Study by Twelve Surgeons", Proc. 6<sup>th</sup> IEEE International Symposium on Mixed and Augmented Reality, Nara, Japan, 2007, pp. 1-4.  
 [7] L. T. De Paolis, M. Pulimeno, M. Lapresa, A. Perrone, and G. Aloisio, "Advanced Visualization System Based on Distance Measurement for an Accurate Laparoscopy Surgery", Proc. Joint Virtual Reality Conf. of EGVE - ICAT - EuroVR, Lyon, France, 2009, pp. 17-18.  
 [8] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab, "Contextual Anatomic Mimesis: Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality", IEEE Proc. Int. Symposium on Mixed and Augmented Reality, Nara, Japan, 2007.  
 [9] S. Nicolau, A. Garcia, X. Pennec, L. Soler, X. Buy, A. Gangi, N. Ayache, and J. Marescaux, "An augmented reality system for liver thermal ablation: Design and evaluation on clinical cases", Elsevier, 2009.  
 [10] LiverPlanner, <http://liverplanner.icg.tu-graz.ac.at>  
 [11] B. Reitingner, A. Bornik, R. Beichel, G. Werkgartner, and E. Sorantin, "Tools for augmented reality based liver resection planning", Proceedings of the SPIE Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display, pages 88-99, San Diego, February 2004.  
 [12] L. Maier-Hein, A. Tekbas, A. Seitel, et al., "In vivo accuracy assessment of a needlebased navigation system for CT-guided radiofrequency ablation of the liver", Medical Physics, 2008, Vol. 35, No. 12, 5385-5396, 0094-2405.  
 [13] T. Stüdeli, D. Kalkofen, P. Risholm, et al., "Visualization tool for improved accuracy in needle placement during percutaneous radiofrequency ablation of liver tumors", Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling, Pts 1 and 2, Vol. 6918, B9180-B9180, 0277-786X.  
 [14] T. S. Yoo, "Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis", A K Peters, Ltd, 2004.  
 [15] Mimics Medical Imaging Software, Materialise Group, <http://www.materialise.com/>  
 [16] 3D Slicer, <http://www.slicer.org>  
 [17] ParaView, <http://www.paraview.org>  
 [18] OsiriX Imaging Software, <http://www.osirix-viewer.com>  
 [19] NDI Polaris Viera, <http://www.ndigital.com>  
 [20] K. Clearya, L. Ibanez, and S. Ranjan, "IGSTK: a software toolkit for image-guided surgery applications", Conference on Computer Aided Radiology and Surgery, Chicago, USA, 2004.  
 [21] L. Weiss, E. Grundmann, J. Torhorst, et al., "ZIB Structure Prediction Pipeline: Hematogenous metastatic patterns incolonic carcinoma: an analysis of 1541 necropsies", Journal of Pathology, vol. 150, pp. 195-203, (1986)  
 [22] Liver Cancer Study Group of Japan, "Primary liver cancer in Japan: clinicopathologic features and results of surgical treatment", Ann Surg, 211, (1990)  
 [23] D. Nagorney, J. Van Heerden, D. Ilstrup, et al., "Primary hepatic malignancy: surgical management and determinants of survival", Surgery, 1989, vol. 106, pp. 740-748.

## ***Increased Cognitive Load in Resolution of Problems Caused by Human Error on New Aircrafts***

Edgard Thomas Martins  
 Depto de Design- CAA  
 Universidade Federal de Pernambuco  
 Recife- Pernambuco - Brasil  
 edgardpiloto@gmail.com

**Abstract-** The flaws in the commitment of decision-making in emergency situations and the lack of perception related to all elements associated with a given situation in a short space of time indicate, often, lack of situational awareness. Automation always surprises the crews and often prevents them from understanding the extent of this technology that is very common in aircraft units with a high degree of automation. These facts are discussed in a subtle way by aircraft drivers who can not do it openly, as it might create an impression of professional self-worthlessness (self-deprecation). This leads to common questions like: What is happening now? What will be the next step of automated systems? This type of doubt would be inadmissible in older aircraft because the pilot of those machines works as an extension of the plane. This scenario contributes to emotional disorders and a growing hidden problem in the aeronautical field. These unexpected automation surprises reflect a complete misunderstanding or even the misinformation of the users. It also reveals their inability and limitations to overcome these new situations that were not foreseen by the aircraft designers. Our studies showed a different scenario when the accident is correlated with systemic variables. It has identified the problems or errors that contribute to the fact that drivers are unable to act properly. These vectors, when they come together, may generate eventually a temporary incompetence of the pilot due to limited capacity or lack of training in the appropriateness of automation in aircraft or even, the worst alternative, due to a personal not visible and not detectable non-adaptation to automation. We must also consider in the analysis the inadequate training and many other reasons, so that we can put in right proportion the effective participation or culpability of the pilot in accidents. Our doctoral thesis presents statistical studies that allow us to assert that the emotional and cognitive overload are being increased with automation widely applied in the cockpits of modern aircraft, and also that these new projects do not go hand in hand with the desired cognitive and ergonomic principles.

**Keywords-***Cognitive overload; New technologies; Automation; Human error.*

### I. INTRODUCTION

The emotional stability and physical health of workers on board aircraft are faced with the factors and conditions that enable professionals to carry out their activities and develop normally, despite the fact that these conditions may present themselves to professionals in adverse conditions [1]. The modern history of aviation with its great technological

complexity has pilots as redundant components that integrate embedded controls in modern aircraft. This leads us to say that the value of the worker as a permanent social group in society does not receive, currently, the proper priority. In research on the health of the pilot, there are three major perspectives that have been investigated that influence his stability, as well as the mental and emotional development of the modern airline pilot [2]:

- The previous life of the individual directly tied to experience, age, genetic and physiological vectors,
- The social environment, cultural environment and formal education leading to the final result, manifested by the ability, personality, strength and character and
- The verifiable standards of quality and quantity of life desired, ambition and achievements and its effects.

The Digital technology advances, has changed the shape and size of instruments used for navigation and communication. This has changed the actions of pilots, especially in relation to emergency procedures. There are few studies that correlate the reduction of accidents with the cognitive and technological changes. The increased cognitive load relates to these changes and requires assessment. The benefits presented by new technologies do not erase the mental models built, with hard work, during times of initial training of the aircraft career pilots in flying schools.

The public must be heeded when an aircraft incident or accident becomes part of the news. In search of who or what to blame, the pilot is guilty and immediately appointed as the underlying factors that involve real evidence of the fact they are neglected.

The reading of the *Black-Boxes* notes that 70% to 80% of accidents happen due to human error, or to a string of failures that were related to the human factor [3]. We can mention stress and the failure to fully understand the new procedures related to technological innovations linked to automation. Complex automation interfaces always promote a wide difference in philosophy and procedures for implementation of these types of aircraft, including aircraft that are different even manufactured by the same

manufacturer. In this case, we frequently can identify inadequate training that contributes to the difficulty in understanding procedures by the crews. Accident investigations concluded that the ideal would be to include, in the pilot training, a psychological stage, giving to him the opportunity of self-knowledge, identifying possible "psychological breakdowns" that his biological machine can present that endangers the safety of flight. Would be given, thus, more humane and scientific support to the crew and to everyone else involved with the aerial activity, minimizing factors that cause incidents and accidents. Accident investigators concluded that the ideal situation for pilot training should include a psychological phase [4], giving him or her, the opportunity of self-knowledge, identifying possible "psychological breakdowns" that biological features can present and can endanger the safety of flight. It should be given, thus, more humane and scientific support to the crew and everyone else involved with the aerial activity, reducing factors that can cause incidents and accidents. Accidents do not just happen. They have complex causes that can take days, weeks or even years to develop [5]. However, when lack of attention and / or neglect take place resulting in a crash, we can be most certain there was a series of interactions between the user and the system that created the conditions for that to happen [6]. We understand that human variability and system failures are an integral part of the main sources of human error, causing incidents and accidents. The great human effort required managing and performing actions with the interface as the task of monitoring, the precision in the application of command and maintaining a permanent mental model consistent with the innovations in automation make it vulnerable to many human situations where errors can occur.

The human variability in aviation is a possible component of human error and we can see the consequences of these errors leading to serious damage to aircraft and people. It is not easy, in new aviation, to convey the ability to read the instruments displays. This can conduct to the deficiency and the misunderstanding in monitoring and performing control tasks: lack of motivation, the fact that it is stressful and tiring, and generate failures in control (scope, format and activation), poor training and instructions that are wrong or ambiguous. The mind of the pilot is influenced by cognition and communication components during flight, especially if we observe all information processed and are very critical considering that one is constantly getting this information through their instruments. There is information about altitude, speed and position of one's aircraft and the operation of its hydraulic power systems. If any problem occurs, several lights will light up and warning sounds emerge increasing the volume and type of man-machine communication which can diminish the perception of detail in information that must be processed and administered by the pilot. All this information must be processed by one's brain at the same time as it decides the necessary action in a context of very limited time. There is a limit of information that the brain can deal with which is part of natural human limitation. It can lead to the unusual situation in which,

although the mind is operating normally, the volume of data makes it operate in overload, which may lead to failures and mistakes if we consider this man as a biological machine. Fig. 1 shows the human-machine interaction where difficulties with cognitive and operational perspectives needs and also physical and emotional aspects take place in a human being during the occurrence of system-level of flight.

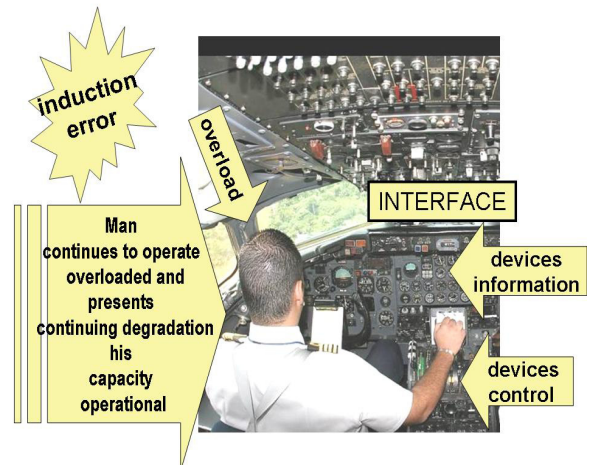


Figure 1 - Diagram of the interaction between man and machine

All situations in which a planned sequence of mental or physical activities fails to achieve its desired outputs are considered as errors [7]. Thus, it is necessary that steps be taken toward reducing the likelihood of occurrence of situations which could cause a problem. The flight safety depends on a significant amount of interpretations made by the pilot in the specific conditions in every moment of the flight. Accidents do not only occur due to pilot error, but also as a result of a poor design of the transmission of information from the external environment, equipment, their instruments, their signs, sounds and different messages. In these considerations, the human agent will always be subject to fatality, which is a factor that can not be neglected. Because of human complexity, it is difficult to convince, in a generic way, people with merely causal explanations. Further analysis of the problem will always end with the identification of a human error, which was probably originated in the design phase, at the manufacturing stage, or given simply as a result of an "act of God". Aeronautical activities, designing human-machine systems becomes very necessary to characterize and classify human error. Human activities have always been confronted with the cognitive system.

## II. CONTEXTUALIZATION

On the result of the causality of accidents, we must consider the human contribution to accidents, distinguishing between active failures and latent failures due to the immediate adverse effect of the system aspect. The main

feature of this component is that it is present within the process of construction of an accident long before declaring the event like an accident, being introduced by higher hierarchical levels as designers, responsible for maintenance and personnel management. We can always guarantee, with respect to organizational accidents, that the layers of defenses, that are the protective barriers, were constructed to prevent the occurrence of natural or man-made disasters. This statement is derived from the design philosophy that treats the defense in depth. In Fig. 2, based on the model "Swiss Cheese" [8], a fail of defense of an accident may occur as a Swiss cheese with "holes", which mean "latent failures" that sometimes began the construction of an accident long before the event. In certain circumstances, such failures (holes) can align themselves and then, the accident happens. An accident is a succession of failures.

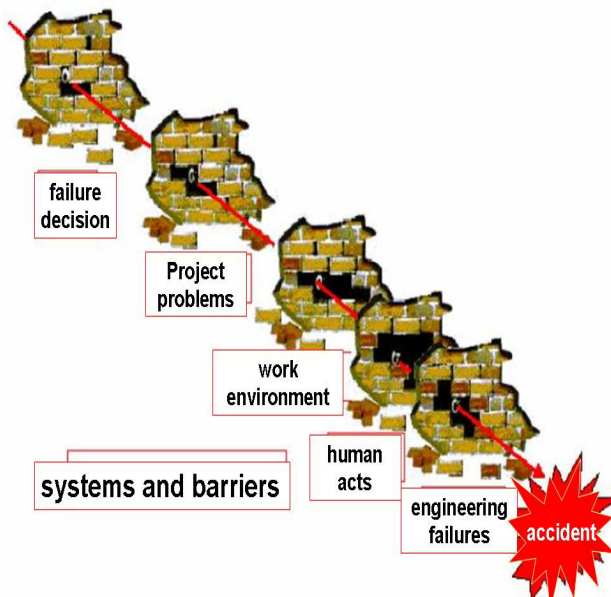


Figure 2- Latent failures, based on the "Swiss Cheese Model" of Reason.

When these barriers are destroyed or are flawed or become vulnerable, the accident occurs. In this fact, that is called latent failure.

### III. CURRENT ACTIVITIES PURSUED BY THE PILOT AND INCREASED COGNITIVE LOAD

The following factors are an integral part of cognitive activity in the pilot: fatigue, body rhythm and rest, sleep and its disorders, the circadian cycle and its changes, the G-force and acceleration of gravity, the physiological demands in high-altitude, night-time take-offs and the problem of false illusion of climbing. But, other physiological demands are

placed by the aviators. It is suggested that specific studies must be made for each type of aircraft and workplace, with the aim of contributing to the reduction of incidents arising from causes so predictable, yet so little studied. We must also give priority to airmen scientists that have produced these studies in physiology and occupational medicine, since the literature is scarce about indicating the need for further work in this direction. Human cognition refers to mental processes involved in thinking and their use. It is a multidisciplinary area of interest includes cognitive psychology, psychobiology, philosophy, anthropology, linguistics and artificial intelligence as a means to better understand how people perceive, learn, remember and how people think, because will lead to a much broader understanding of human behavior.

Cognition is not presented as an isolated entity, being composed of a number of other components, such as mental imagery, attention, consciousness, perception, memory, language, problem solving, creativity, decision making, reasoning, cognitive changes during development throughout life, human intelligence, artificial intelligence and various other aspects of human thought [9].

The procedures of flying an aircraft involve observation and reaction to events that take place inside the cabin of flight and the environment outside the aircraft [10]. The pilot is required to use information that is perceived in order to take decisions and actions to ensure the safe path of the aircraft all the time. Thus, full use of the cognitive processes becomes dominant so that a pilot can achieve full success with the task of flying the "heavier than air."

With the advent of automated inclusion of artifacts in the cabin of flight that assist the pilot in charge of controlling the aircraft, provide a great load of information that must be processed in a very short space of time, when we consider the rapidity with which changes occur, an approach that cover the human being as an individual is strongly need. Rather, the approach should include their cognition in relation to all these artifacts and other workers who share that workspace [11].

### IV. CONDITIONS FOR THE DEPLOYMENT OF THE TASKS LEADING TO ACCIDENTS.

A strong component that creates stress and fatigue of pilots, referred to the design of protection, detection and effective handling of fire coming from electrical short circuit on board, is sometimes encountered as tragically happened on the Swissair Airlines flight 111, near Nova Scotia on September 2, 1998. The staff of the Federal Aviation Administration (FAA), responsible for human factors research and modern automated interfaces [12], reports a situation exacerbated by the widespread use an electrical product and a potentially dangerous wire on aircrafts, called "Kapton".

If a person has to deal with an outbreak of fire, coming from an electrical source at home, the first thing he would do is disconnect the electrical power switch for the fuses. But this option is not available on aircraft like the Boeing B777 and new Airbus. The aviation industry is not

adequately addressing the problem of electrical fire in flight and is trying to deal recklessly [13]. The high rate of procedural error associated with cognitive errors, in the automation age, suggests that the projects in aviation have ergonomic flaws. In addition, it has been related that the current generation of jet transport aircraft, used on airlines, like the Airbus A320, A330, A340, Boeing B777, MD11 and the new A380 (see Fig. 3), that are virtually "not flyable" without electricity. We can mention an older generation, such as the Douglas DC9 and the Boeing 737.



Figure 3- The automated Airbus A380 cockpit totally dependent on electricity (Photos from the Collection of the author)

Another factor in pushing the pilots that causes emotional fatigue and stress is the reduction of the cockpit crew to just two. The next generation of large transport planes four engines (600 passengers) shows a relatively complex operation and has only two humans in the cockpit. The flight operation is performed by these two pilots, including emergency procedures, which should be monitored or re-checked. This is only possible in a three-crew cockpit or cockpit of a very simple operation. According to the FAA, the only cockpit with two pilots that meets these criteria is the cabin of the old DC9-30 and the MD11 series (see Fig. 4). The current generation of aircraft from Boeing and Airbus do not fit these criteria, particularly with respect to engine fire during the flight and in-flight electrical fire.



Figure 4 - The cockpit of the old Douglas MD-11 (Photos from the Collection of the author)

The science of combining humans with machines requires close attention to the interfaces that will put these components (human-machine) working properly. The deep study of humans shows their ability to instinctively assess and treat a situation in a dynamic scenario. A good ergonomic design project recognizes that humans are fallible and not very suitable for monitoring tasks. A properly designed machine (such as a computer) can be excellent in monitoring tasks. This work of monitoring and the increasing the amount of information invariably creates a cognitive and emotional overload and can result in fatigue and stress.

According to a group of ergonomic studies from FAA [14] in the United States this scenario is hardly considered by the management of aviation companies and, more seriously the manufacturers, gradually, introduce further informations on the displays of Glass cockpits. These new projects always determine some physiological, emotional and cognitive impact on the pilots.

The accident records of official institutes such as the NTSB (National Transportation Safety Bureau, USA) and CENIPA (Central Research and Prevention of Accidents, Brazil) show that some difficulties in the operation, maintenance or training aircraft, which could affect flight safety are not being rapidly and systematically passed on to crews worldwide. These professionals of aviation may also not be unaware of the particular circumstances involved in relevant accidents and incidents, which makes the dissemination of experiences very precarious.

One of the myths about the impact of automation on human performance: "while investment in automation increases, less investment is needed in human skill" [15]. In fact, many experiments showed that the progressive automation creates new demands for knowledge, and greater, skills in humans. Investigations of the FAA [16] announced that aviation companies have reported institutional problems existing in the nature and the complexity of automated flight platforms. This results in additional knowledge requirements for pilots on how to work subsystems and automated methods differently. Studies showed the industry of aviation introduced the

complexities of automated platforms flight inducing pilots to develop mental models about overly simplified or erroneous system operation. This applies, particularly, on the logic of the transition from manual operation mode to operation in automatic mode [17]. The process of performing normal training teaches only how to control the automated systems in normal but do not teach entirely how to manage different situations that the pilots will eventually be able to find. This is a very serious situation that can be proved through many aviation investigation reports that registered the pilots not knowing what to do, after some computers decisions taken, in emergencies situations [18]. VARIG (Brazilian Air lines), for example, until recently, had no Boeing 777 simulators where pilots could simulate the emergence loss of automated systems what should be done, at list, twice a month, following the example of Singapore Airlines.

According to FAA [19], investigations showed incidents where pilots have had trouble to perform, successfully, a particular level of automation. The pilots, in some of these situations, took long delays in trying to accomplish the task through automation, rather than trying to, alternatively, find other means to accomplish their flight management objectives. Under these circumstances, that the new system is more vulnerable to sustaining the performance and the confidence. This is shaking the binomial Human-Automation compounded with a progression of confusion and misunderstanding. The qualification program presumes it is important for crews to be prepared to deal with normal situations, to deal with success and with the probable. The history of aviation shows and teaches that a specific emergency situation, if it has not happen, will certainly happen.

#### V. NEW FOCUS FOR PROCEED AN SYSTEMIC EVALUATION IN THE PERFORMANCE OF THE PILOTS

Evaluating performance errors, and crew training qualifications, procedures, operations, and regulations, allows them to understand the components that contribute to errors. At first sight, the errors of the pilots can easily be identified, and it can be postulated that many of these errors are predictable and are induced by one or more factors related to the project, training, procedures, policies, or the job. The most difficult task is centered on these errors and promoting a corrective action before the occurrence of a potentially dangerous situation.

The FAA team, which deals with human factors [20], believes it is necessary to improve the ability of aircraft manufacturers and aviation companies in detecting and eliminating the features of a project, that create predictable errors. The regulations and criteria for approval today do not include the detailed project evaluation from a flight deck in order to contribute in reducing pilot errors and performance problems that lead to human errors and accidents. Neither the appropriate criteria nor the methods or tools exist for designers or for those responsible for regulations to use

them to conduct such assessments. Changes must be made in the criteria, standards, methods, processes and tools used in the design and certification. Accidents like the crash of the Airbus A320 of the AirInter (a France aviation company) near Strasbourg provide evidence of deficiencies in the project.

This accident highlights the weaknesses in several areas, particularly when the potential for seemingly minor features has a significant role in an accident. In this example, inadvertently setting an improper vertical speed may have been an important factor in the accident because of the similarities in the flight path angle and the vertical speed in the way as are registered in the FCU (Flight Control Unit).

This issue was raised during the approval process of certification and it was believed that the warnings of the flight mode and the PFD (Primary Flight Display-display basic flight information) would compensate for any confusion caused by exposure of the FCU, and that pilots would use appropriate procedures to monitor the path of the vertical plane, away from land, and energy state. This assessment was incorrect. Under current standards, assessments of cognitive load of pilots to develop potential errors and their consequences are not evaluated. Besides, the FAA seeks to analyze the errors of pilots, a means of identifying and removing preventively future design errors that lead to problems and their consequences. This posture is essential for future evaluations of jobs in aircraft crews.

Identify projects that could lead to pilot error, prematurely, in the stages of manufacture and certification process will allow corrective actions in stages that have viable cost to correct or modify with lower impact on the production schedule. Additionally, looking at the human side, this reduces unnecessary loss of life.

#### VI. CONCLUSION AND FUTURE WORK

We developed a study focusing on the guilt of pilots in accidents when preparing our thesis. In fact, the official records of aircraft accidents blame the participation of the pilots like a large contributive factor in these events.

Modifying this scenario is very difficult in the short term, but we can see as the results of our study, which the root causes of human participation, the possibility of changing this situation. The cognitive factor has high participation in the origins of the problems (42% of all accidents found on our search). If we consider other factors, such as lack of usability applied to the ergonomics products, choice of inappropriate materials and poor design, for example, this percentage is even higher.

Time is a factor to consider. This generates a substantial change in the statistical findings of contributive factors and culpability on accidents. The last consideration on this process, as relevant and true, somewhat later, must be visible solutions. In aviation, these processes came very slowly, because everything is wildly tested and involves many people and institutions. The criteria adopted by the official organizations responsible for investigation in aviation accidents do not provide alternatives that allow a clearer view of the problems that are consequence of

cognitive or other problems that have originate from ergonomic factors. We must also consider that some of these criteria cause the possibility of bringing impotence of the pilot to act on certain circumstances. The immediate result is a streamlining of the culpability in the accident that invariably falls on the human factor as a single cause or a contributing factor. Many errors are classified as only "pilot incapacitation" or "navigational error". Our research shows that there is a misunderstanding and a need to distinguish disability and pilot incapacitation (because of inadequate training) or even navigational error.

Our thesis has produced a comprehensive list of accidents and a database that allows extracting the ergonomic, systemic and emotional factors that contribute to aircraft accidents. These records do not correlate nor fall into stereotypes or patterns. These patterns are structured by the system itself as the accident records are being deployed. We developed a computer system to build a way for managing a database called the Aviation Accident Database. The data collected for implementing the database were from the main international entities for registration and prevention of aircraft accidents as the NTSB (USA), CAA (Canada), ZAA (New Zealand) and CENIPA (Brazil). This system analyses each accident and determines the direction and the convergence of its group focused, instantly deployed according to their characteristics, assigning it as a default, if the conditions already exist prior to grouping. Otherwise, the system starts formatting a new profile of an accident [21].

This feature allows the system to determine a second type of group, reporting details of the accident, which could help point to evidence of origin of the errors. Especially for those accidents that have relation with a cognitive vector. Our study showed different scenarios when the accidents are correlated with multiple variables. This possibility, of course, is due to the ability of Aviation DataBase system, which allows the referred type of analysis.

It is necessary to identify accurately the problems or errors that contribute to the pilots making it impossible to act properly. These problems could point, eventually, to an temporary incompetence of the pilot due to limited capacity or lack of training appropriateness of automation in aircraft. We must also consider many other reasons that can alleviate the effective participation or culpability of the pilot. Addressing these problems to a systemic view expands the frontiers of research and prevention of aircraft accidents.

This system has the purpose of correlating a large number of variables. In this case, the data collected converges to the casualties of accidents involving aircraft, and so, can greatly aid the realization of scientific cognitive studies or applications on training aviation schools or even in aviation companies [22]. This large database could be used in the prevention of aircraft accidents allowing reaching other conclusions that would result in equally important ways to improve air safety and save lives.

## VII. REFERENCES

- [1] C. Eugenio, *Automação no cockpit das aeronaves: um precioso auxílio à operação aérea ou um fator de aumento de complexidade no ambiente profissional dos pilotos*, 1rd ed., vol. 1. São Paulo: Brasil, 2011, pp. 34-35.
- [2] E. Henriqson, "Coordination as a Distributed Cognitive Phenomena Situated in Aircraft Cockpits- Aviation in Focus," "A Coordenação Como Um Fenômeno Cognitivo Distribuído e Situado em Cockpits de Aeronaves," UFRGS edit., Porto Alegre, vol. 12, pp. 58 –76, Dez. 2010.
- [3] Federal Aviation Administration, "DOT/FAA/AM-10/13, Office of Aerospace Medicine, Causes of General Aviation Accidents and Incidents: Analysis Using NASA Aviation Safety Reporting System Data, Washington DC, U.S. Department of Transportation press, Sept. 2010.
- [4] S. Dekker, "Illusions of explanation- A critical essay on error classification," *The International Journal of Aviation Psychology*, New Jersey, vol. 13, pp. 95-106, Sept. 2003.
- [5] J. Reason, *Human Error*, 2nd ed., Cambridge: U.K., University Press, 1990, pp.92-93.
- [6] J. Rasmussen, "A taxonomy for describing human malfunction in industrial installations," *The Journal of Occupational Accidents*, vol. 4, Jul. 1982, pp. 311-333.
- [7] J. Reason, *The Smart Human*, 5rd ed., Cambridge: U.K., University Press, 2008, p. 347.
- [8] J. Reason, *The Smart Human*, 5rd ed., Cambridge: U.K., University Press, 2008, p. 347
- [9] R. Sternberg, *Cognitive psychology*, Porto Alegre: Brasil, Ed Artmed, 2000.
- [10] R.G. Green and M. Frenbard, *Human Factors for Pilots*. Avebury: England, Technical Aldershot, 1993.
- [11] Federal Aviation Administration- FAA, "The Interface Between Flightcrews and Modern Flight Deck Systems," Federal Aviation, Administration Human Factors Team Report, May 2005. pp-34-35.
- [12] Federal Aviation Administration- FAA, "The Interface Between Flightcrews and Modern Flight Deck Systems," Federal Aviation, Administration Human Factors Team Report, May 2005, pp-177-178.
- [13] Federal Aviation Administration- FAA, "The Interface Between Flightcrews and Modern Flight Deck Systems," Federal Aviation, Administration Human Factors Team Report, May 2005, pp 113-116.
- [14] Federal Aviation Administration- FAA, "The Interface Between Flightcrews and Modern Flight Deck Systems," Federal Aviation, Administration Human Factors Team Report, May 2005, pp 22-28.
- [15] Federal Aviation Administration, "Human Error Analysis of Asrts Reports: Altitude Deviations in Advanced Technology Aircraft," Federal Aviation Administration, Human Factors Team Report, 1992, 1996, 2002.
- [16] National Transportation Safety Board, "Airline Service Quality Performance- User Manual- ASQP," Washington, DC, Bureau of Transportation Statistics, U.S. Department of Transportation press, Oct. 2011, pp. 75-77.

- [17] National Transportation Safety Board, "Airline Service Quality Performance- User Manual- ASQP," Washington, DC, Bureau of Transportation Statistics, U.S. Department of Transportation press, Oct. 2011, pp. 123-128.
- [18] National Transportation Safety Board, "C.F.R.- 234 Airline Service Quality Performance Reports," Washington, DC, Bureau of Transportation Statistics, Research and Innovative Technology Administration (RITA), U.S. Department of Transportation press, Oct. 2011, pp. 45-61.
- [19] Federal Aviation Administration- FAA, "Human Error Analysis of Accidents Report," Federal Aviation Administration- Human Factors Team Report, 2010, pp. 201-206.
- [20] Federal Aviation Administration- FAA, "Human Error Analysis of Accidents Report," Federal Aviation Administration- Human Factors Team Report, 2010, pp. 108-115.
- [21] E. Martins, "Ergonomics in Aviation: A critical study of the causal responsibility of pilots in accidents," "Ergonomia na Aviação: Um estudo crítico da responsabilidade dos pilotos na causalidade dos acidentes," Msc. Monography, Universidade Federal de Pernambuco, Pernambuco: Brasil, Mar. 2007, pp. 285-298.
- [22] E. Martins, "Study of the implications for health and work in the operationalization and the aeronaut embedded in modern aircraft in the man-machines interactive process complex," "Estudo das implicações na saúde e na operacionalização e no trabalho do aeronauta embarcado em modernas aeronaves no processo interativo homem-máquinas complexas," thesis, Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Pernambuco:Brasil, Aug. 2010, pp. 567-612.



# Using Social Media for Collaborative Learning in Higher Education: A Case Study

Na Li, Sandy El Helou, Denis Gillet  
 École Polytechnique Fédérale de Lausanne (EPFL)  
 1015 Lausanne, Switzerland  
 {na.li, sandy.elhelou, denis.gillet}@epfl.ch

**Abstract**—This paper investigates the acceptability of using social media for collaborative learning in the context of higher education. A social media platform, *Graasp*, is used to support students' learning activities in a project-based course. An evaluation of *Graasp* regarding its usefulness as a collaboration platform, a knowledge management site, and a gadget container, was conducted with the course participants. Quantitative and qualitative assessment methods used in the evaluation, as well as the main findings are presented.

**Keywords**—social media; collaborative learning; knowledge management; E-Learning 2.0

## I. INTRODUCTION

With the rise of social media, Web users have become co-producers of social content rather than passive information consumers. Next to its wide usage for social interactions among young people, social media is also increasingly used to support learning activities. Many efforts have been made to incorporate social media into students' overall learning ecology, which leads to the emergence of E-Learning 2.0. The concept of "E-Learning 2.0" [1] refers to the adoption of social media in learning or education where learners are empowered to create and organize their own learning activities.

In practice, higher education institutions are still primarily relying on traditional learning management systems (LMS) that do not fully capitalize on the potential of social media for enabling participation in global learning networks, collaboration and social networking [2] [3]. According to a previous study by Clark et al. [4], while Web 2.0 participatory technologies have become an essential part of young learners' daily lives, very few learners are taking full advantage of these technologies to support their learning processes. Similarly, Greenhow & Robelia also argued in their study [5] that students do not perceive a connection between their online activities and institutional learning. Overall, the studies suggest that the potential benefits of using social media to create learner-centered education systems need to be further exploited and well understood by learners.

In this paper, we investigate the added value of using social media for collaborative learning in the context of higher education. A social media platform, namely *Graasp* (formerly *Graasp* <http://graasp.epfl.ch>), is introduced to a number of undergraduate students, and then used in one of their courses to support project-based collaborative learning.

The acceptability and expectation of such social media supported learning solutions is examined through a user study. The rest of the paper is organized as follows. Section II discusses the existing attempts of integrating social media into students' learning experience. Section III presents the structure and features of *Graasp*. The context of the user study is described in Section IV. *Graasp*'s acceptability in sustaining collaborative learning is evaluated and analyzed in Section V. Finally Section VI concludes the paper.

## II. RELATED WORK

Compared to traditional learning management systems that provide few opportunities for learners to develop and maintain their own learning activities, learning platforms based on social media paradigms place the control of learning into the hands of learners themselves [6]. A growing number of research efforts have been made to support teaching and learning using a variety of social media tools. For example, Silius et al. [7] developed a social networking site for college students, aiming at enhancing both collaborative study and social interaction. Their research reveals that making social media tools a part of traditional learning is attractive to students and can motivate their participation in the learning process. In other similar studies, a social bookmarking tool [8], a blogging platform [9], or wiki software [10] have been used to engage students in collaborative projects and encourages creating, editing, and sharing content. The research suggests that, it is promising to adopt social media in the context of learning as it promotes collective knowledge generation and encourages active user participation. However, the application of social media to enhance learning solutions is still in its early phase and needs to be further explored. In our previous work [6], we have presented the social media platform *Graasp*, and investigated its potential role for sustaining collaborative learning activities through scenarios. In this paper, a real-world experiment has been conducted to evaluate the added value of using *Graasp* for students' collaborative work in a project-based course.

## III. A SOCIAL MEDIA PLATFORM: *GRAASP*

*Graasp* is a Web 2.0 application developed at Swiss Federal Institute of Technology in Lausanne (EPFL in French). It can serve simultaneously as an aggregation, contextualization, discussion, and networking platform, a shared asset repository, or an activity management system. The structure of *Graasp* relies on the extension of the 3A

interaction model [11], which is intended for designing and describing social and collaborative learning environments. The 3A model consists of three main constructs or entities: **Actors** represent entities capable of initiating an event in a collaborative environment, such as regular users or virtual agents. Actors create personal or collaboration spaces where they conduct personal and group **Activities** to reach specific objectives. In each of these activities, actors can take different roles with different set of rights. In addition, actors produce, edit, share and annotate **Assets** in order to meet activity objectives. Assets can consist of simple text files, RSS feeds, videos or audio files. The 3A model is extended to incorporate a fourth entity, namely **Applications**, to account for widgets or gadgets [12] that can be added and executed within activity spaces. **SALT** is an acronym adopted to describe the social media actions of Sharing, Assessing (i.e. rating and commenting), Linking, and Tagging that can be performed by actors on assets and applications in the context of activity spaces. These actions are believed to encourage contribution, collaboration, and reflection [11]. In *Graasp*, Actors, Activities, Assets, and Applications are mapped to People, Activity Spaces, Resources, and Apps respectively. *Graasp*'s main features that facilitate collaborative learning are discussed hereafter.

#### A. Collaboration within Spaces

The design of *Graasp* follows a bottom-up flexible approach that releases hierarchical and constraining default structures when it comes to managing joint projects. Instead of having a top-level administrator in control of all project spaces, everything is managed at the space level. Both tutors and students are entitled to create activity spaces. Spaces owners are free to choose between hierarchical or completely flat structures. In a flat-structured space, every member shares equal rights so that no one acts as a supervisor who superintends the learning process. As an example, a team of students is able to create a project space where all members share, discuss, organize learning resources (using tags or sub-spaces), and collaboratively coordinate the project's activities. Additionally, *Graasp* also enables top-down hierarchical structured spaces where members can take different roles. For instance, tutors can create a course space to define the course milestones, post learning materials, and organize the learning activities. In order to keep users aware of the ongoing activities, *Graasp* enables them to subscribe to RSS feeds within a space to trace changes.

#### B. Collective Knowledge Generation and Management

Unlike traditional learning management systems where learners are only allowed to access the curriculum given by the tutor, *Graasp* empowers students to create, aggregate, share, and organize the learning resources by themselves. *Graasp* offers typical social media features such as tags, ratings, comments, wikis, and bookmarks. Tags enable users to classify their collections in the ways they find useful, and also facilitate building a folksonomy in the learning community [13]. Such bottom-up classification is particularly helpful for efficient search and recommendation since user-defined tags make it easier to discover relevant

items. With respect to ratings and comments, they provide an easy way for users to express their preferences and thus help evaluating the quality of user-generated content from the community perspective. As far as wiki is concerned, it enables users to co-create social content and cooperatively work towards common goals. It is worth mentioning that *Graasp* provides a bookmarking feature, "*GraaspIt!*", with which users are able to grab and link external Web pages into their *Graasp* spaces. This feature facilitates knowledge aggregation from a variety of public sources.

#### C. Learning-Oriented Gadgets

Not only does *Graasp* serve as a knowledge management and collaboration platform, but also as a gadget container. Gadgets can run and communicate within *Graasp*. This capability reinforces the learning experience because it enables useful learning-oriented tools to be added and launched during the learning process. Different collections of gadgets can be associated to different spaces, making the aggregation contextual. Thanks to this feature, *Graasp*'s provided functionalities are made flexible and extendable. For instance, in a project space, students can add a calendar gadget configured with a series of milestones and deadlines. Other gadgets such as notepad, to-do list, and learning plan application could be useful as well.

#### D. Privacy Control Scheme

Since *Graasp* provides a relatively open learning environment, there is a clear need for effective privacy control mechanisms that protect against unauthorized access to social data. Instead of adopting complicated privacy management schemes that are difficult for users to cope with, the privacy settings are maintained at the space level. Based on its purpose and its owner's choice, a space can be public, closed, or hidden. Public spaces are globally visible and allow every user to join. Closed and hidden ones are only accessible upon explicit invitations. Hidden spaces are not searchable and they are only visible to space members. Closed and hidden spaces are especially useful when students want to carry out their peer-based projects without being disturbed by others or feeling that they are "observed" by the tutor.

Within a specific space, users are allowed to take different roles: owner, editor, and viewer. Each role is associated with a set of rights allowing users to perform diverse actions such as moderating the space, adding new assets in the space, commenting, rating, tagging, bookmarking, and so on. Only space owners are entitled to assign roles to other space members. Assigning different roles in a collaborative space makes users aware of their duties and gives them the opportunity to concretely collaborate by being allowed to perform specific actions. Finally, the access permissions of assets and gadgets inherit from the space they belong to.

### IV. USING *GRAASP* FOR COLLABORATIVE LEARNING

To examine the acceptability and user satisfaction of *Graasp* in terms of supporting collaborative learning, it was used as a collaborative work platform in a project-based

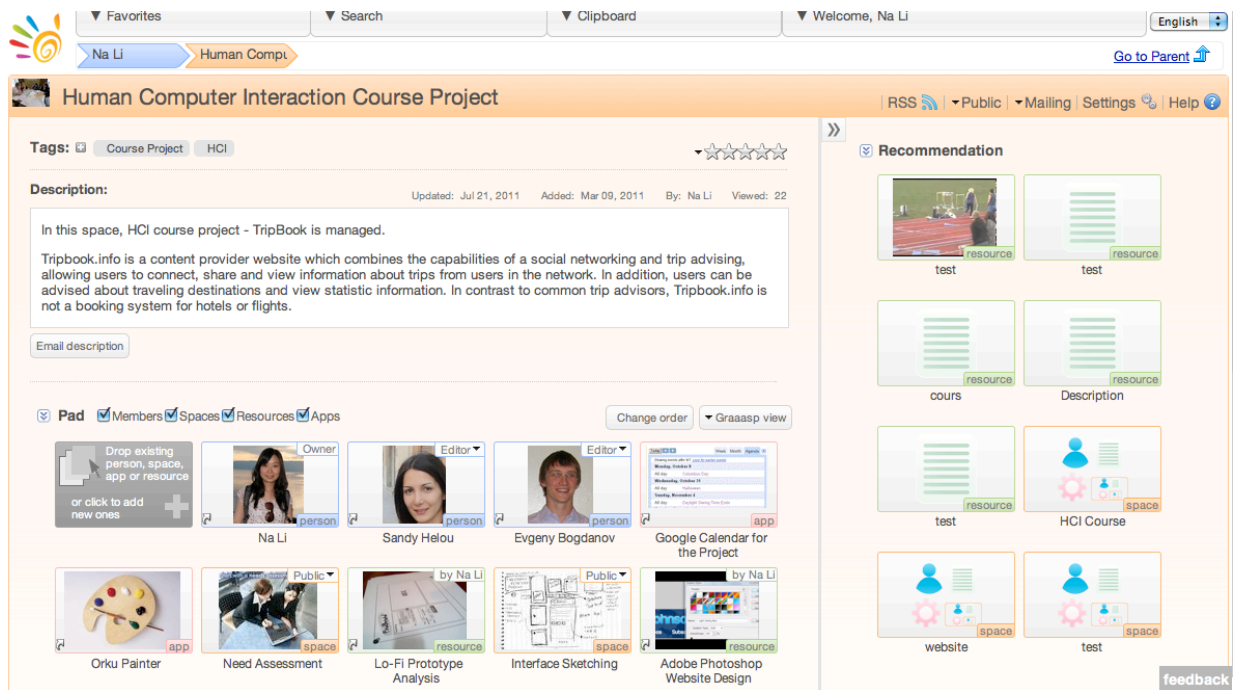


Figure 1. A project space in *Graasp*

course of *Human Computer Interaction* offered at Tongji University in China. 28 undergraduate students were involved in the course and they were divided into 8 teams. Each team was required to accomplish a group project. *Graasp* was introduced to the students at the beginning of the course. Since it is an intensive course, the students' teamwork was limited to two 4-hours face-to-face sessions and a 30 minutes' final presentation spread over a three-week period. As a result, the total usage of *Graasp* was not expected to be very high.

After a brief introduction of *Graasp*, students were shown how they could create their project spaces, share resources with each other, play different roles in the project, and work with different gadgets. An example of a project space is illustrated in Fig. 1. The "*Human Computer Interaction Course Project*" space is shown on the left side of the user interface, and the recommended items are shown on the right side. The space members, sub-spaces, posted assets (resource), and added gadgets (app) are displayed in the "*Pad*" of the space.

During the course, each group of students was entitled to create their project space (as shown in Fig. 1), and invite all the team members to join. They then chose a privacy type for their space, and assigned different permissions to the team members depending on their different roles. Within the space, students could co-edit the wiki and make discussions through comments. Course resources were added into the space either through direct creation within *Graasp* or by grabbing external data using "*Graaspl!*". Space members could also create sub-spaces in order to organize and structure their collaborative work if necessary. A few

gadgets (e.g., mockup gadget and project management gadget) that were believed to promote the learning process were provided to the students to work with. Students could also "grasp" more gadgets from online gadget repositories such as iGoogle (<http://www.google.com/ig/directory?type=gadgets>).

## V. EVALUATION AND ANALYSIS

To investigate the acceptability of *Graasp* as a collaborative learning platform, a user study was conducted with the students participating in the course described in Section IV. The evaluation methodology and main findings are addressed in this section.

### A. Evaluation Methodology

The evaluation consists of two parts: a quantitative assessment relying on students' action log throughout the period of the course, and a qualitative assessment through a user questionnaire distributed after the course.

In the first part of the evaluation, students' activities within *Graasp* are analyzed in the following aspects: the number of entities (people profiles, activity spaces, assets, and apps or gadgets) created, the number of invitations sent to join a space, the number of tags, comments, ratings, and wiki generated, and the use of privacy control feature. The objective is to examine the general usage of *Graasp* and students' online interactions during the course.

In the second part of the evaluation, an online questionnaire was distributed to the course participants after the final presentation of their teamwork. The user questionnaire is composed of Likert-scale questions [14]

with 5-point preference scale (strongly disagree, disagree, neutral, agree, and strongly agree), multiple choice questions and open questions. The questions can be grouped into the following categories: the general usefulness of *Graasp*, the usefulness of *Graasp* as a collaboration platform, the usefulness of *Graasp* as a knowledge management system, the usefulness of learning-oriented gadgets, and the user satisfaction regarding the privacy control scheme. The user questionnaire is summarized in Table I. In addition, we also used the Desirability Toolkit approach [15], where the study participants were asked to select 5 adjectives that can best describe their personal experience with *Graasp*, from a list of 118 words. Through this approach, we got a depth of understanding and authenticity in participants' experience and opinions on *Graasp*.

TABLE I. USER QUESTIONNAIRE

<b>I</b>	<p><b>General usefulness</b></p> <p>Generally speaking, I find <i>Graasp</i> useful for my teamwork.</p> <p>I think <i>Graasp</i> is useful for the following reasons (Easy sharing, Project management, Resource organization, Resource aggregation, and Gadgets integration)</p> <p>Using <i>Graasp</i> improves my motivation for carrying out my teamwork.</p> <p>I predict that I would frequently use <i>Graasp</i> in the future.</p>
<b>II</b>	<p><b>Usefulness as a collaboration platform</b></p> <p>I feel comfortable to do teamwork with my classmates in a project space.</p> <p>I find it useful to add an item in a space and share it with space members.</p> <p>Have you organized your people connections using spaces? Why?</p>
<b>III</b>	<p><b>Usefulness as a knowledge management system</b></p> <p>I like to use spaces to organize my resources.</p> <p>I think it is useful to aggregate different resources into <i>Graasp</i> using "<i>Graaspl!</i>".</p> <p>Have you used the wiki? Why?</p> <p>Have you used tags? Why?</p>
<b>IV</b>	<p><b>Usefulness of learning-oriented gadgets</b></p> <p>I think the following gadgets are helpful for my learning or working activities (Project management gadget, Mockup gadget, and Other gadgets I found).</p> <p>I think it is a good idea to integrate gadgets into my learning or working process.</p> <p>What other gadgets will be useful for your collaborative learning and teamwork?</p>
<b>V</b>	<p><b>User satisfaction of privacy control scheme</b></p> <p>I am satisfied with having control over the privacy settings (public, closed, hidden) of my spaces and profile.</p> <p>I feel it is necessary to assign different permission (owner, editor, and viewer) to others over my resources.</p> <p>I would like to have an option to specify more detailed permissions, such as "who can tag on me", "who can comment on my space", and so on.</p>

**B. Study Participants**

The user study was conducted with 28 undergraduate students, who were the intended audience of the *Graasp*

system. All participants were frequent Web users who were familiar with social media platforms and Web 2.0 technologies like blogging, wiki, rating, tagging, and bookmarking. However, most of them did not have much experience of using online learning systems. Instead, the most common tools used for teamwork were instant messenger and email. A few students claimed to use shared calendars for collaborative projects.

**C. Results and Discussion**

The quantitative results extracted from students' action log and the qualitative feedback collected through user questionnaires are discussed in this section. The number of items created throughout the course period and the proportion of students using the corresponding feature are illustrated in Fig. 2 and Fig. 3 respectively.

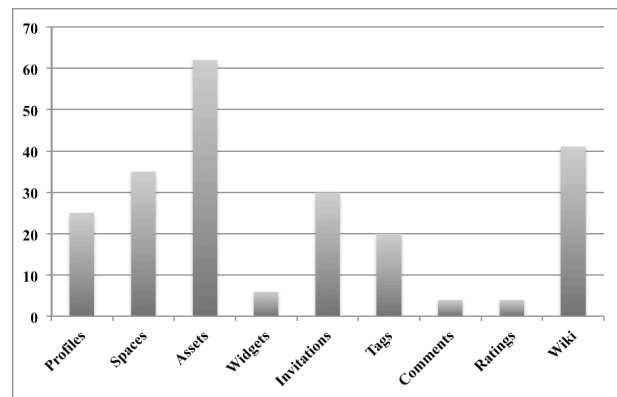


Figure 2. Numbers of items created during the course period

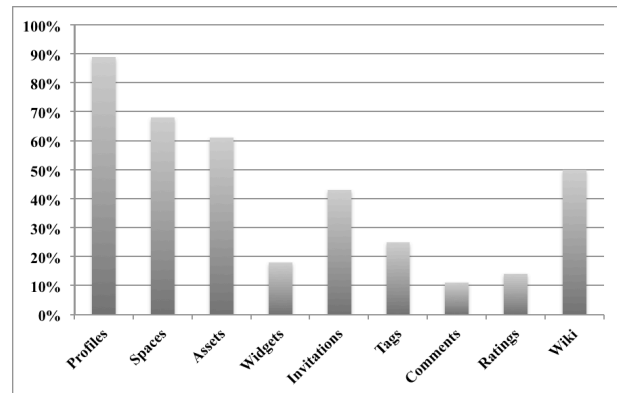


Figure 3. Proportion of students using the features

As shown in the figures, 25 out of 28 students registered with *Graasp*, 35 new spaces were created by 19 students (25 home spaces were also created by default when the users registered), 62 assets were posted by 17 students, and only 6 gadgets were added by 5 students in total. The results indicate that the overall usage of *Graasp* was not high, which matches our initial expectation since students were given only a few hours to carry out their teamwork. In contrast to the other entities, the reason why a small number of gadgets

were created might be the lack of course relevant gadgets in the online gadgets repositories. Within 35 spaces, 30 invitations were sent, and 22 of them were accepted. This suggests that although the general usage of *Graasp* was not high, some active users utilized space as a place to collaborate with each other.

Regarding the use of typical Web 2.0 features, 20 tags, 41 wiki, 4 comments, and 4 ratings were generated (as shown in Fig. 2). Students later in the user questionnaire explained that they thought tags were helpful to search for relevant items, and also facilitate describing and classifying content. Students also pointed out that wiki played an important role in providing basic information of an item and defining the learning context. This is consistent with the fact that the use of tags and wiki was relatively active (as shown in Fig. 3). The reason behind the low usage of comments and ratings was explained to be the fact that there was not sufficient social data in the current system to evaluate.

As for the use of privacy control feature, among 60 spaces including 25 default home spaces, 19 of them were set to public, 17 ones were closed, and 24 ones were hidden. The diversity of privacy settings confirms the students' answers in the questionnaire that it is necessary to set different privacy levels to spaces depending on what a specific space is used for.

In addition to the quantitative assessment through the log of students' actions, user questionnaires were also collected and analyzed after the course. All the 28 user questionnaires were successfully completed. The results and findings are presented hereafter.

With respect to *Graasp's* general usefulness, 64% participants considered *Graasp* useful as a platform for sharing and organizing resources, 46% of them perceived it as an adequate place to collaboratively manage their projects, and 46% of them recognized its usefulness as a system aggregating content from various sources. Slightly over a half of the students (52%) confirmed that *Graasp* improved their motivation for carrying out their teamwork. A reason why the result is not satisfying enough is due to the technical problem of slow network connection in the campus. Also, the students pointed out that they felt less motivated since there were not sufficient peer connections besides the rest of their classmates attending the course. Furthermore, the lack of social resources in the current system somehow disappointed them as well. To solve this problem, the capability of exchanging social resources and personal contacts with popular social media platforms could be a possible direction. The interoperability with such platforms needs to be further explored in *Graasp*.

The rest of the questionnaire was intended to investigate students' satisfaction of *Graasp's* usefulness in different aspects. The qualitative results shown in Fig. 4 reveal that the majority of the students were satisfied with using *Graasp* for those purposes. More specifically, regarding the usefulness of *Graasp* as a collaboration platform, 57% of the students expressed their preference for carrying out teamwork within project spaces. 74% of them thought it useful to share items with others using spaces. 17 participants claimed that they organized their people

connections by placing them into different spaces. Students pointed out that it was an effective way to classify personal contacts depending on different context. However, 11 students never used this feature. The reason given was that there were not yet many contacts to classify, and all their contacts were their classmates. In short, the role of *Graasp* in supporting collaborative work is quite satisfying. Most people thought it comfortable to undertake teamwork using *Graasp*.

From the perspective of supporting collective knowledge management, 59% of the participants considered it convenient to structure and organize resources using different spaces, and a slightly higher proportion of them (67%) affirmed the usefulness of aggregating resources using "*GraaspIt!*" feature. Moreover, 17 out of 28 students stated that tags helped them to describe and classify content easily, while the remaining said that there was no need to use tags due to the lack of data existing in the system.

When asked whether learning-oriented gadgets can enhance their learning experience or not, 63% of the students confirmed that the integration of gadgets into their learning process was helpful. It is worth mentioning that there is a huge difference between the percentage of users' preference (63%) and their real usage in practice (18%). This implies that although users are hoping to use gadgets for collaborative learning, it is difficult for them to find useful gadgets intended for learning activities in online repositories. Therefore, more efforts are needed to develop project-management and learning-oriented gadgets that could comply with learners' learning requirements. In the questionnaire, students also proposed a few potential gadgets that could help them achieve their learning goals, such as online chatting tool, learning plan application, to-do list with reminder, and so on.

As far as the privacy control feature is concerned, 74% of the students were satisfied with having control over the privacy levels of spaces and user profiles. Granting different permissions to others over their personal resources was perceived as an essential feature as well. In the current *Graasp* system, only three types of permissions are allowed: owner, editor, and viewer. It was suggested by the participants that more fine-grained permission rules should be added. Users should be capable of defining their own

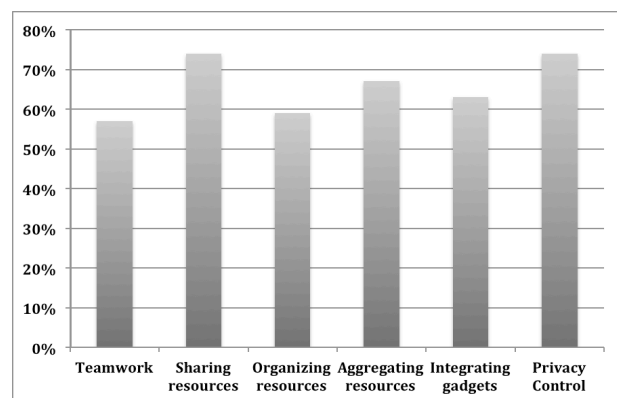


Figure 4. Students' satisfaction of *Graasp's* usefulness

permission rules, such as “who can tag on me”, “who can comment on my space”, and “who can link my resources by drag and drop”. This requires further evaluation, and raises the challenges of a trade-off between more fine-grained choice and an increased system complexity.

Finally, we asked the participants to pick 5 adjectives that closely matched their personal reactions to *Graasp* from a list of positive and negative words. A word cloud, showing the frequency of the selected adjectives, is presented in Fig. 5. Among a variety of words, the most frequently picked ones were accessible, personal, trustworthy, reliable, and slow. The overall assessment is quite encouraging with a few negative opinions like slowness. One should note that the performance problem is mainly due to the limited network bandwidth in the campus and the cross-country data transmission.



Figure 5. A word cloud based on frequency of selected adjectives

## VI. CONCLUDING REMARKS

This paper discussed the usage of social media in learning contexts. A social media platform, namely *Graasp*, simultaneously offering as a bottom-up project management platform, a resource repository, a collaboration site, and a gadget container, is presented. The evaluation of its acceptability among students taking a HCI course is also discussed. Despite a technical problem of bandwidth that causes slow server response, results show that students were satisfied with using *Graasp* to enhance knowledge management and collaboration. Organizing content using subspaces or tags was less exploited due to the lack of data in the platform. Students mentioned that to be able to use *Graasp* outside the scope of the course, it would be useful to be able to import contacts from other popular sites. Future work will focus on interoperability efforts to ease the exchange of resources and contacts with other social media platforms based on standards such as OpenSocial, already used for inter-gadget communications.

Within the framework of another research project, *Graasp* is now being deployed and increasingly used by students from the Universities of Geneva and Fribourg. Large-scale evaluations involving courses that span over a longer period of time will be conducted to further examine the added value of using *Graasp* for collaborative learning.

## ACKNOWLEDGEMENT

The research work described in this paper is partially funded through the ROLE Integrated Project; part of the Seventh Framework Program for Research and Technological Development (FP7) of the European Union in Information and Communication Technologies.

## REFERENCES

- [1] C. Safran, D. Helic, and C. Gütl, “E-Learning practices and Web 2.0,” Proc. International Conference on Interactive Computer Aided Learning, vol. 1, pp. 1-8, 2007.
- [2] N. Dabbagh and A. Kitsantas, “Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning,” Internet and Higher Education (2011), in press.
- [3] C. McLoughlin and M. Lee, “Personalised and self regulated learning in the Web 2.0 era: International exemplars of innovative pedagogy using social software,” Australasian Journal of Educational Technology, vol. 26, pp. 28-43, 2010.
- [4] W. Clark, K. Logan, R. Luckin, A. Mee, and M. Oliver, “Beyond Web 2.0: mapping the technology landscapes of young learners,” Journal of Computer Assisted Learning, vol. 25, pp. 56-69, 2009.
- [5] C. Greenhow and B. Robelia, “Informal learning and identity formation in online social networks,” Learning, Media and Technology, vol. 34, pp. 119-140, 2009.
- [6] N. Li, C. Ullrich, S. El Helou, and D. Gillet, “Using social software for teamwork and collaborative project management in higher education,” Proc. International Conference on Web-based Learning, vol. 6483, pp. 161-170, 2010.
- [7] K. Silius, T. Miilumäki, J. Huhtamäki, T. Tebest, J. Meriläinen, and S. Pohjolainen, “Students’ motivations for social media enhanced studying and learning,” Knowledge Management & E-Learning: An International Journal, vol. 2, pp. 51-67, 2010.
- [8] T. M. Farwell and R. D. Waters, “Exploring the use of social bookmarking technology in education: an analysis of students’ experiences using a course-specific Delicious.com account,” Journal of Online Learning and Teaching, vol. 6, pp. 398-408, 2010.
- [9] D. Rosen and C. Nelson, “Web 2.0: a new generation of learners and education,” Computers in the Schools, vol. 25, pp. 211-225, 2008.
- [10] S. Hazari, A. North, and D. Moreland, “Investigating pedagogical value of wiki technology,” Journal of Information Systems Education, vol. 20, pp. 187-198, 2009.
- [11] S. El Helou, N. Li, and D. Gillet, “The 3A interaction model: towards bridging the gap between formal and informal learning,” Proc. International Conference on Advances in Computer-Human Interactions, pp. 179-184, 2010.
- [12] Gadgets Specification, <http://code.google.com/apis/gadgets/docs/spec.html>, retrieved Sep. 2011.
- [13] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, and J. Riedl, “Tagging, communities, vocabulary, evolution,” Proc. Anniversary Conference on Computer Supported Cooperative Work, pp. 181-190, 2006.
- [14] R. Likert, “A technique for the measurement of attitudes,” Archives of Psychology, vol. 140, 1932.
- [15] J. Benedek and T. Miner, “Measuring desirability: new methods for evaluating desirability in a usability lab setting,” Redmond, WA: Microsoft Corporation, 2002.

## AlgoPath: A New Way of Learning Algorithmic

Estelle Perrin

CReSTIC

IFTS – University of Reims  
Charleville-Mézières, France  
estelle.perrin@univ-reims.fr

Sébastien Linck

IFTS – University of Reims  
Charleville-Mézières, France  
sebastien.linck@univ-reims.fr

Frédéric Danesi

DINCCS

Charleville-Mézières, France  
frederic.danesi@dinccs.com

**Abstract—** This paper presents a new way of learning algorithmic: AlgoPath is a virtual world in which variables are represented by 3D figures carrying a backpack and the sequence of instructions is represented by a stone path. The interface of AlgoPath helps students to avoid common mistakes. The world of AlgoPath gives them a mental representation of algorithms. Students are more prone to learn because AlgoPath changes the level of difficulty. They can forget the off-putting syntax and grammar of algorithmics. AlgoPath is ludic and students feel they are more playing than learning.

**Keywords-** 3D-based training; education; algorithmics; ludic teaching

### I. INTRODUCTION

Learning how to create algorithms can sometimes be frustrating for students in the urge of using computers. We now live in a world where numericals are everywhere: playing is numerical, reading is numerical, and socializing can be numerical. Asking students to handle a sheet of paper and a pencil in order to learn the basic concepts of programming is kind of out of phase. Bringing back numericals in the process of learning can be perceived reassuring for students and can increase their interest and their concentration.

But numericals cannot solve everything. They have to be attractive enough or they can be classified as being as boring as a sheet of paper and a pencil. On the contrary, students can learn while playing with a good numerical game. Nowadays games are virtual and teenagers agree that the most important characteristics in a game are those that contributed to its realism [1].

In our university, we do have students that have to learn algorithmics although they don't have an extended scientific background. One of our under-graduate courses is dedicated to websites – how to design, how to add on modules, how to write for the web, etc. Some of the students are more familiar with communications while others are more familiar with informatics. But they all have to attend a course dedicated to algorithmics. The students whose mind is not shaped to rationalize and to think in a way that fits logics, sometimes have difficulties to catch a good mental representation of an algorithm [2] and sometimes are not motivated to do so. As hard teachers can keep trying to explain, a variable or a loop, for example, is and remains abstract and some students just do not conceptualize it. As a consequence, these students

will not even bother learning the syntax of algorithms and they will fail without a doubt. Algorithmics is just something they are not motivated for. So they need something attractive to help them learn and avoid errors. Using computer games is one means to encourage learners to learn. Even if game-based learning is not the most efficient learning method per se, games enhance motivation and increase students' interest in the subject matter [3]. Indeed, there have already been quite a lot of research projects towards the development of software games for education that aim at increasing the students' motivation and engagement while they learn [4][5].

In Section II we will review visual programming dedicated to algorithmics and we will show that their interfaces are traditional while the game-based environment is more enjoyable for the users than a traditional environment [6]. In Section III, we will show that algorithms can be represented as a virtual world in which 3D figures walk along a path. In Section IV, we will briefly show the implementation of an example from scratch and we will finally conclude and reveal future work in Section V.

### II. OVERVIEW

In this section, we study various graphical algorithms and programming editors. All of them have advantages but also different disadvantages. We will focus on two characteristics of these editors: (1) the appearance of the editor (see Section “A”) and (2) the representation of the programming code or algorithm (see Section “B”).

The software studied is designed to help students learning algorithmics whatever their age. Some of the software, mostly those dedicated to algorithm editors, are developed by standalone or amateur developers. The others are created by researchers or programmers. For example, Algoris is a software created by an amateur who makes small programs for pleasure [7]. He developed Algoris to help young and older students to learn algorithmics with a visual method. Scratch was developed by the Lifelong Kindergarten Research Group at the MIT Media Laboratory [8]. It was designed to introduce concepts to children in mathematics and computer science [9].

#### A. Appearance

The colored visualization of a program or algorithm, whatever the kind of code it is, is an important point to help students to understand.

Colors can be used as an attention-getting element in program interfaces. Algoris and Scratch follow this method by using a colored enhancement in the text and the representation of their algorithms [10]. Both editors have rounded and colored interfaces but although they look like identical, they have different aims. The first one is to learn algorithmic to create real codes. The second one is to simplify the creation of software. In first case the aim is to learn, in the second the aim is to learn unaware of doing so.

Scratch looks like a game. Its instructions are represented by boxes that fit like Lego bricks. These bricks are classified in categories (for example to move or to speak), which complicates the task of learning the program. Algoris has a form of circuit like a flowchart. A ball goes through boxes connected by pipes. It is a helpful representation to visualize what is inside the loop and what is done or not. Alice2 enhances the visualization thanks to a programming environment designed by building 3D virtual worlds [11].

Not all learning algorithm or programming software is beautifully designed.

But colors can be distracting too. With dark colors and shapes, LARP [12] and AlgoBox [13] seem less practical, stricter, and so more serious. Algobox is not graphic. The code is represented by text and is well made up. It follows the structure of an algorithm. LARP constructs algorithms in a flowchart form. It is pretty visual, but the user has to know the representations of every element. However its flowchart is not attractive to the eye. Jeliot, a java editor is divided into two parts [14]. The first part is a text editor (for programming purposes) with highlighted syntax keywords. The second part, called the “theater”, is hidden by a blue curtain during the drafting stage of the code. At the beginning of the execution the curtain opens and reveals a four-part window. These parts represent the element classes of the program (method, expression evaluation, constant, instance, and array). The execution modifies adequately each part at each step of the program. The advantage of this software stops there, since the user has to type in the whole source of his algorithm.

Alvis is apparently a fairly classic program [15]. When you open it you first see a text editor. A whiteboard with a toolbox is placed next to it: this complementarity makes Alvis interesting. Alvis makes a parallel between the source code and the representation of the elements in memory: boxes are for variables and boxes containing boxes are for arrays.

One important standard of a good learning program is the appearance. Scratch combines the most attractive one and a playful display of the content of the variables. Alice2 makes a good evolution in representation by introducing 3D and can be seen as a gateway to the game world.

For some of the programs, there is a real time link between the graphical and the textual representation of the algorithm. It means that if the user changes the source, the representation of the memory will be changed instantaneously. Scratch, Jeliot [16], and Alvis [17] are the only ones which propose this automatically, but it is also proposed on-demand in the other programs.

## B. Code Structure

AlgoBox and Scratch variable declarations are very simple. The user just clicks on a button called “Declare new variable” or “New variable”. There is no graphical representation of a variable in AlgoBox but Scratch illustrates the concept of a variable with a box whose value is adjustable with a sliding scale. With Algoris or LARP, there is no declaration. Whenever a variable is needed in a program, it is set without your having to give a type. With Alvis, the drag and drop of a new variable from the menu to the editor opens a window assignment. To learn the notion of type and variables, Alvis seems better.

The Scratch conditional instructions seem the simplest and most visual implementation. The flowchart representation is quite visual on this point. One can easily understand how the algorithm works and what instructions are to be executed.

In Algobox, there is a button to create conditional instructions. Everything is explained in the initialization window, where specifications can be added at request.

With Alvis, there is no a graphical representation of the conditional instruction, but we are guided step by step to build it. Alvis helps the user in the establishment of a simple conditional instruction, but you have to modify the code to create a compound one. Finally with Algoris there are two boxes for “if” and one for “if / else”.

The representation in Scratch or LARP is quite visual, it so it is easy to understand the course of a conditional instruction. Furthermore the definition of the conditions of the conditional instructions is designed like a puzzle.

The representation of loops in Scratch is interesting. They are represented in the form of pliers in which the instructions to be repeated are set. The problem is that the syntax is not usual, which is a problem for learning algorithms. Yet the visual aspect of the loops makes it easy to understand that they are the instructions to be repeated.

LARP is based on the usual algorithmic syntax, it's easy to understand what and each loop how does, but with flowcharts it is not easy to recognize what kind of loop it is.

With AlgoBox, there is still nothing graphic, but the initialization window is suitable and all the necessary explanations are given. For the loop, the variable that is chosen will be used as a counter. The starting and final values can be chosen by the user but not the increment which is necessarily of one.

In LARP, the code generation of a loop is well explained thanks to the initialization window where you define all the parameters available. The arrangement of loop representation in Scratch and Algoris helps to understand the sequence of loops.

All this algorithm and code editors, although very different by their languages or representations, are helpful for the user as they provide a representation of the code and a better highlighted syntax.



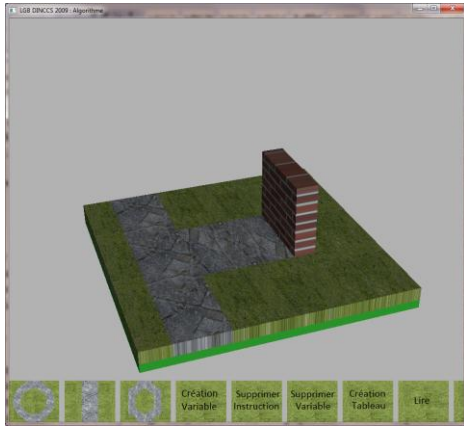


Figure 1. An empty algorithm in AlgoPath.

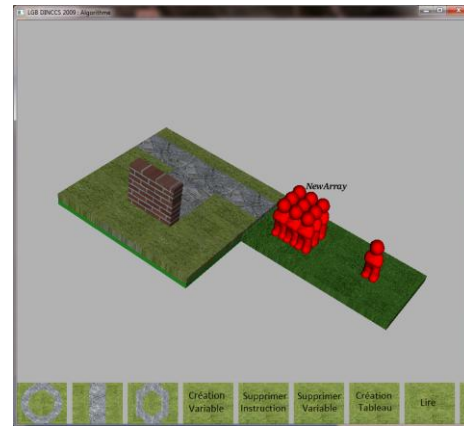


Figure 2. Creation of an array

Graphical representation or model is a good methodology to understand and interpret algorithms [18]. For some of them, the learning of a programming language is simplified by creating a new language between code and algorithm.

### III. ALGOPATH

The main advantages of AlgoPath are:

- AlgoPath is attractive. Students face a virtual world which is like a game environment.
- AlgoPath helps them learn because it gives a mental representation of an abstract concept that is the algorithm.
- AlgoPath helps them avoid some common mistakes. Students can forget the off-putting syntax and grammar.
- AlgoPath lets them focus on the solving of the problem they have to automate without syntax or grammar errors. You will see that, at the end, a 3D figure just has to carry the proper value.
- AlgoPath changes the level of difficulty. It is easier to learn and to understand [19].

In AlgoPath, algorithms are a world reduced to a stone path lined with grass and trees (see Figure 1). In such a state, the algorithm represented by AlgoPath simply does nothing.

In the next sections, we will show how variables and basic instructions are represented and the last section will be dedicated to parameters passed by value or by reference.

#### A. AlgoPath Variables

In AlgoPath, a variable is a 3D figure (see Figure 2.). The user always begins with virtual declarations of variables because she/he cannot do anything else. Actually, no instruction can be added if no variable has been created. To do so, she/he just has to create a new 3D figure and give it a name. Then a new 3D figure stands in a platform next to the stone path lined with grass and folds its arms across its chest. Why does the 3D figure seem unoccupied?

It appears so because it has not been set yet. As expected, several instructions are available and are described in the next sections:

- The assignment.
- The conditional instructions.
- The loop instructions.

An array can also be created by the user of AlgoPath. For teaching purposes, only one dimensional and two dimensional arrays are available in AlgoPath. In that case, the array is shown by multiple 3D figures standing right next to the others.

#### B. AlgoPath Assignment

The assignment sets or re-sets the value associated to a variable. In AlgoPath, the user starts with specifying the right-value of the assignment with the keyboard and the mouse and then links it to a 3D figure. During this step, she/he just has to click on the appropriate 3D figure the software shows. An ordinary mistake our students do is to invert the right value with the left-value of an assignment. This cannot happen in AlgoPath. They also cannot use an unset variable in a right-value because the software only shows them the set variables during the specification of the right-value.

Once the variable has been set, the 3D figure carries a backpack. It also walks along the stone path (see Figure 3.). The difference between the green platform and the one with the stone path is that the green platform contains all the 3D figures that can be used in an instruction. The fact that a 3D figure standing in the green platform carries a backpack reminds the user that the associate variable has already been set.

When assigning an array, the user clicks on the corresponding 3D figure but this interaction is not sufficient. She/he also needs to choose which slot of the array is assigned. During this stage, the software shows the user multiple 3D figures: one for each slot of the array. The names of these 3D figures are the name of the array concatenated to their position in the array (zero stands for the first slot).

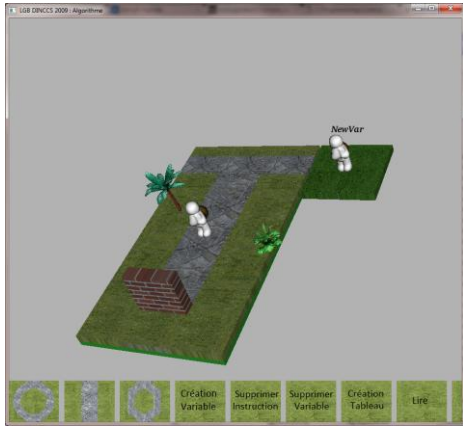


Figure 3. A set variable and its assignment.

At any time, the user can double-click on the 3D figure and see the right value of the assignment.

*C. AlgoPath Conditional Instructions*

AlgoPath conditional instructions change the stone path (see Figure 4). Instead of being a straight line, it splits into two paths. There is a traffic light at the intersection of the paths which stands for the condition of the conditional instructions.

If the algorithm could be executed and the condition turned out to be true, the traffic light would turn out to be green. If the condition turned out to be false, the traffic light would turn out to be red. That is why, one path is covered with a green bush, the other with a red bush.

After the creation of a conditional instruction, the user has to click on both the green and the red bushes. By doing so, she/he will have to complete the unconditional instructions and the conditional instructions. At each step, a new empty stone path is shown and has to evolve. If she/he creates a new variable when the shown stone path is one of the conditional or unconditional instructions, then the associated 3D figure is only available in the conditional or unconditional instructions.

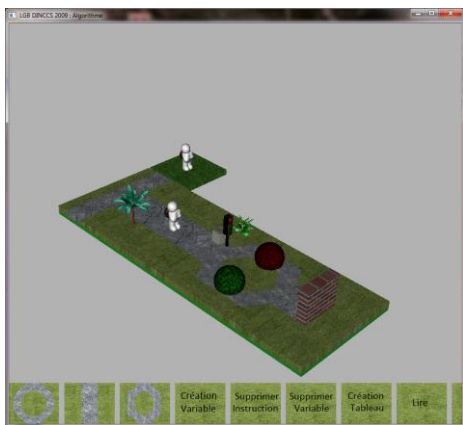


Figure 4. An AlgoPath conditional instruction.

It means that the user will no longer see it when she/he leaves the stone paths of the conditional instruction. This specification helps the students learn the notion of the scope of a variable. They are able to see it before any program is run.

As for the right value of an assignment, the user can see the condition (by double-clicking on the traffic light), the conditional instructions (by double-clicking on the green bush), and the unconditional instructions (by double-clicking on the red bush) of a conditional instruction at any time.

*D. AlgoPath Loop Instructions*

As an AlgoPath conditional instruction, an AlgoPath loop instruction changes the stone path (see Figure 5). It becomes a two-entry traffic circle. At one of these two entries stands a traffic light. The position of the traffic light specifies the type of the loop. If the traffic light is at the beginning of the loop, the instructions of the loop may never be executed (it is a while loop). If the traffic light is at the end of the loop, a walker could be free to enter the loop. It means that the instructions of the loop could be executed at least once (it is a repeat loop).

After the creation of a loop instruction, the user has to double-click on the bush of the traffic circle. By doing so, she/he has to complete the instructions of the loop. Then a new empty stone path is shown and has to evolve. During this step, if a new variable has been created, the related 3D figure is no longer known when the user leaves the loop stone path. Again, the student faces the notion of the scope of a variable.

*E. AlgoPath Input and Output*

AlgoPath has two more instructions. One if the future user's algorithm wants to see the value of a variable (a slate is added in the stone path) and another if the future user's algorithm has to set a variable (the 3D figure shows an open book).

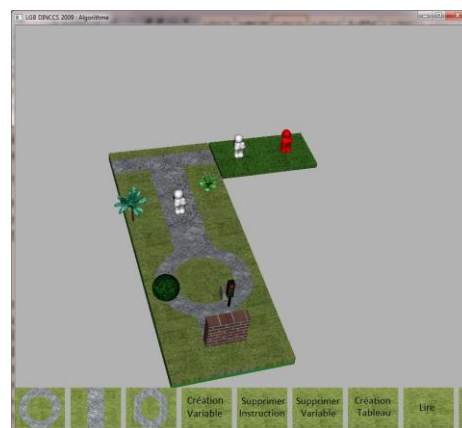


Figure 5. An AlgoPath loop instruction.

### F. Passing data to procedures and functions

In programming, there are two ways of passing data to procedures and functions: either by value (in that case, parameter in the call of a procedure or a function does not share memory with the parameter in the definition of a procedure or a function) or by reference (in that case, they share memory). In AlgoPath, students are able to visualize this process. As passing data by value passes a copy of the data for processing and as a data is represented by a 3D figure carrying a backpack, parameters by value are visualized by two 3D figures standing in front of each other, the first one showing its backpack to the second one, and the second one carrying its own backpack. On the contrary, as passing data by reference saves changes made to data and passes those changes back to the calling algorithm, parameters by reference are visualized by two 3D figures facing each other and holding the same backpack (see Figure 6).

The call of a function or a procedure is represented by a forest. Multiple 3D figures stand at the edge of the forest. They are the parameters of the function or the procedure. For educational purposes, a function never has parameters by reference but obviously passes a result to the calling algorithm. This result is represented by a 3D figure that stands at the exit of the forest. On the contrary, a procedure does not need such a 3D figure but may have parameters by reference.

When the user creates a new function or procedure, she/he must:

- Specify the name of the procedure or the function. With that name, AlgoPath knows if the user will have to create the definition of the function or the procedure in the next step.
- Specify the parameters. If the function or the procedure has already been defined, AlgoPath shows the user the parameters of the called algorithm. The user just has to drag and drop the correct 3D figures in front of the calling parameters. If not, AlgoPath expects the user to drag and drop the correct 3D figures and to choose the correct 3D figure that must stand in front of them to choose a parameter by reference or by value. If she/he chooses a 3D figure with its hands wide open then the parameter is by reference. If she/he chooses a 3D figure which carries its own backpack then the parameter is by value. If it is a function to be defined, a special calling parameter is automatically created with its hands wide open. It is the result of the function.
- Specify the instructions. This step is optional if the procedure or the function has already been defined. If not, the user has to double-click on the forest and a new empty stone path appears. In that case, the platform next to the stone path is already full of 3D figures that correspond to the calling parameters. If the calling parameter is by value then the called parameter has given it its value. So the relative 3D figure carries a backpack. If the calling parameter is

by reference then the called parameter gives it its state. If the called parameter is set, the relative 3D figure carries a backpack. If it is not, it does not. The calling parameter representing the result of a function always folds its arms across its chest at the beginning of the definition of a function and means that it has to be set. AlgoPath does not allow the user to go back to the called algorithm if there is one 3D figure with its arms across its chest left. At any time, the user is able to change the type of a parameter either by double-clicking on the 3D figures at the edge of the forest or by double-clicking on the platform next to the stone path.

### IV. IMPLEMENTATION

AlgoPath has been developed in C++ language using the VTK (Visualization Tool Kit) library. 3D figures have been made in clay and digitized into STL files using the 3DSom Software (see Figure 7). They have been smoothed and enlarged using GOM Inspect and 3D Reshaper.

A more complete example of an algorithm can be seen in Figure 8.

### V. CONCLUSION AND FUTURE WORK

This paper shows that it is possible to translate algorithms into a virtual world where 3D figures are data storage locations, backpacks carried by 3D figures are values and, bushes and forests are a set of instructions and paths are the way instructions are supposed to flow.

We have also shown that AlgoPath can help students avoid common mistakes (such as using a variable in an expression before it is set or forgetting to assign a parameter passed by reference) through an easy and intuitive interface. It helps them learn because AlgoPath gives them a way to have a mental representation of algorithms. AlgoPath is ludic because it is a world close to games and students feel they are playing more than learning. As a consequence, it is easier to learn algorithmic with AlgoPath.

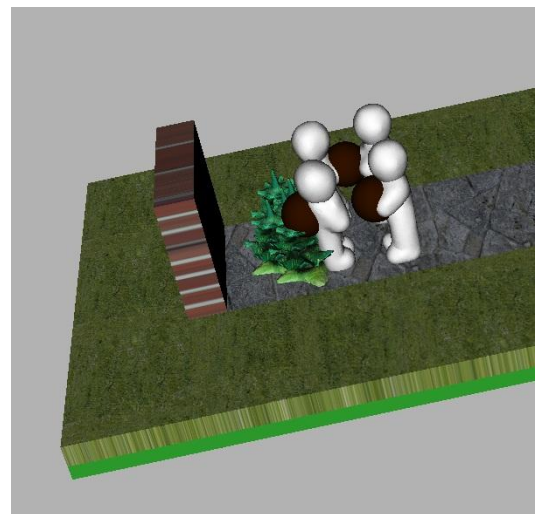


Figure 6. Parameters by reference (in back) and by value (in front).

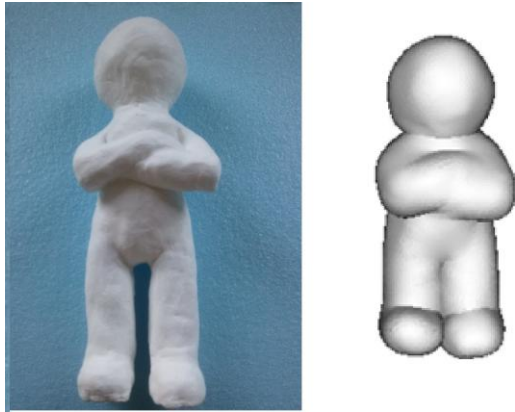


Figure 7. Clay 3D figure and the virtual one.

However, some improvements need to be done. In particular the results of the ongoing user study may point inconsistencies in the interface or in the process.

Anyhow this process must be extended by adding an execution that could be simulated by a bus taking on board every 3D figure along the stone path and delivering them at the end of the algorithm with the correct values.

AlgoPath could turn into a more serious game if the wanted algorithm was chosen by the teacher and was a data of AlgoPath so AlgoPath could help the students find the correct algorithm.

#### ACKNOWLEDGMENT

We thank Marie François and Estelle Bodart for their valuable help during the study and the implementation of AlgoPath.



Figure 8. A more complete example

#### REFERENCES

[1] [http://www.mediaawareness.ca/english/resources/research\\_documents/studies/video\\_games/vgc\\_preferences.cfm](http://www.mediaawareness.ca/english/resources/research_documents/studies/video_games/vgc_preferences.cfm)

[2] Ronit Ben-Bassat Levy, Mordechai Ben-Ari, Pekka A. Uronen, “The Jeliot 2000 program animation system”, *Computers & Education*, vol. 40 (1), pp. 15–21, 2003

[3] Druckman D., “The Educational Effectiveness of Interactive Games. Simulation and gaming across disciplines and

cultures: ISAGA at a watershed”, Thousand Oaks: Sage, pp. 178 – 187, 1995

[4] Alan Amory, Kevin Naicker, Jackie Vincent, Claudia Adams, “Computer games as a learning resource”, *ED-MEDIA, ED-TELECOM '98*, world conference on education multimedia and educational telecommunications, vol. 1, pp. 50–55, 1998.

[5] Cristina Conati, Xiaoming Zhou, “Modeling students’ emotions from cognitive appraisal in educational games”, *Intelligent Tutoring Systems*, pp. 944–954, 2002

[6] Viswanath Venkatesh, “Creation of favorable user perceptions: exploring the role of intrinsic motivation”, *MIS Quarterly*, vol. 23, pp. 239-261, 1999

[7] <http://www.mikesoft.fr/math/Algoris/Logiciel%20algoris.htm>

[8] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, B. Silverman, Y. Kafai, “Scratch: Programming for All”, *Communications of the ACM*, vol. 52 (11), November 2009

[9] M. Resnick, “All I Really Need to Know (About Creative Thinking) I Learned (By Studying How Children Learn) in Kindergarten”, *Proceedings of the SIGCHI Conference on Creativity and Cognition*, Washington, D.C., 2007

[10] Norma S. Pribadi, Maria G. Wadlow, Daniel Boyarski, “The use of color in computer interfaces : Preliminary Research”, *Information technology Center, Carnegie Mellon University, Pittsburgh, Pennsylvania*, 1990.

[11] C. Kelleher Dennis and D. Cosgrove and D. Culyba and C. Forlines and J. Pratt and Y Pausch, *Alice2: Programming without Syntax Errors*, 2002

[12] <http://larp.marcolavoie.ca/fr/default.htm>

[13] <http://www.xmlmath.net/algoebox/>

[14] Mordechai Ben-Ari, Roman Bednarik, Ronit Ben-Bassat Levy, Gil Ebel, Andrés Moreno, Niko Myller, Erkki Sutinen, “A decade of research and development on program animation: The Jeliot experience”, *Journal of Visual Languages and Computing*, vol. 22, oct. 2011, pp. 375-384, 2011

[15] Hundhausen, C.D. & Brown, J.L. What You See Is What You Code: A "Radically-Dynamic" Algorithm Visualization Development Model for Novice Learners." *Proceedings IEEE 2005 Symposium on Visual Languages and Human-Centric Computing*. Piscataway, NJ: IEEE.

[16] Andrés Moreno, Niko Myller, Erkki Sutinen, M. Lattu, V. Meisalo, J. Tarhio “A visualization tool as a demonstration aid”, *Computers & Education* 41(2), pp. 133–148, 2008

[17] Christopher D. Hundhausen, Jonathan L. Brown, “What You See Is What You Code: A “live” algorithm development and visualization environment for novice learners”, *Journal of Visual Languages & Computing*, vol. 18 (1), February 2007, pp. 22-47, 2007

[18] Richard E. Mayer, “Different problem-solving competencies established in learning computer programming with and without meaningful models”, *Journal of Educational Psychology*, vol. 67, pp. 725-734, 1975

[19] John Maloney, Kylie Peppler, Yasmin B. Kafai, Mitchel Resnick, Natalie Rusk, “Programming by Choice: Urban Youth Learning Programming with Scratch”, *SIGCSE'08*, March 12–15, Portland, Oregon, USA, 2008

# Alleyoop: Interactive Information Retrieval System with Sketch Manipulations

Hiroaki Tobita

Sony Computer Science Laboratory Paris

Paris, France

tobby@csl.sony.fr

**Abstract**—The Alleyoop system enables users to retrieve information through sketch manipulations. A number of information retrieval systems already enable users to casually search and browse through the Web. These systems are useful for conventional input forms where a user uses a keyboard to input a keyword to a dialog box. However, they are not suitable for pen-based input styles. Users of pen-based computers have to input a query to a fixed dialog box by drawing the query. In contrast, the Alleyoop system is designed for pen-based computing, so users can interactively retrieve information through sketch manipulations. When a user draws a closed curve and a keyword, information nodes related to the keyword are collected automatically inside the closed curve. The user can also create a Venn diagram by continuously drawing closed curves and keywords, and form more complex queries for information retrieval. Moreover, the system allows the user to create a layout by drawing strokes freely, so he/she can set the information nodes on the layout and see them in detail. In this paper, we describe our Alleyoop system and how it can be effectively applied.

**Keywords**—*Information Retrieval; Information Visualization; Sketch and Paint Manipulations; Interactive System.*

## I. INTRODUCTION

As computer hardware and software are continuously improved, it becomes possible to more quickly express a wider variety of information (e.g., images, movies, documents, and Web pages). Also, many types of computers (e.g., augmented reality (AR) systems, ubiquitous computers, and personal digital assistants (PDAs)) have been developed to allow users to naturally interact with the information space. In the future, these diversifications will be advanced rapidly, so natural and simple visualization techniques for information retrieval are needed.

Numerous systems have been designed for information retrieval. Web based search engines (e.g., Google and Yahoo) have become popular for casual use. Combining rough and detail search methods is one important element, since they have to treat huge amounts of information. These systems are based on conventional input methods, so a keyboard is used to input a keyword to a fixed dialog box. On the other hand, pen-based computers such as PDAs and tablet PCs have become popular recently. They are characterized by simple sketch manipulations similar to drawing a picture on paper in the real world. This sketch-based interaction has been developed especially for creative activities [4, 5, 6]. Sketch interaction is not especially useful for communicating details, but is effective for approximate and casual use. In these systems, a user can use the entire

system window as a workspace. Also, sketch based information retrieval systems have been developed. When a user draws a picture, the system finds similar images from the database automatically. However, they do not provide combination between rough and detail search methods. Some systems focus on the relationship between sketch and layout. Although, Maya paint effects allows users to set pseudo-3D objects in the scene using 2D brush strokes, their target is not information retrieval. Thus, as the original applications have not supported information retrieval, users have to use conventional retrieval applications with pen-based input styles.

Therefore, we have been developing a unique information retrieval system based on simple sketch manipulations for pen-based computers [1, 2]. A feature of our design is that all the manipulations required for information retrieval are based on sketch manipulations and that a user can use the entire window space as a search area. Thus, the user can freely use the whole application window as both an input and a search area. In order to retrieve information, the Alleyoop system mainly provides two interactions: rough and detail searches. The rough search is to collect related information nodes. A user of the Alleyoop system first draws a closed curve and then draws a keyword inside the curve. The system automatically recognizes both the curve area and the keyword after the user has drawn them. Information nodes related to the keyword are then collected within the curve area. In addition, by making a continuous series of simple drawings, the user can create a Venn diagram to form a more complex query. The detail search is to see the information nodes in detail. After collecting the nodes, the user creates an original layout by drawing strokes. The information nodes are set onto the layout, so the user can see the nodes in detail. As both search interactions are based on sketch manipulations, the user can interact with an information space freely and easily.

In this paper, we describe a prototype of the Alleyoop system and how it enables users to handle information through interactive drawing. The next section describes the related work of the Alleyoop system. Section 3 describes the system design. The system overview is mentioned in detail in Section 4. In Section 5, we describe the implementation of the system. In Section 6, we discuss the system based on comments made by visitors. Section 7 concludes the paper.

## II. RELATED WORK

Our system focuses on pen-based information retrieval and allows users to retrieve information through creative activities. Here, we describe work related to our concepts such as information visualization and sketch interfaces.

While considerable research has been done to support the use of information visualization [3] for retrieving information, existing visualization systems are not designed to treat all data sets in a common way. Actually, by applying visualization techniques such as zooming and scaling, specialized systems can be designed to enable efficient visualization of huge amounts of data [8, 9, 11, 12], but such systems are not effective for simple visualization of moderate amounts of data. In particular, in the case of 3D layouts, the navigation methods become as important as the layouts as the amount of data increases. Although some systems include different layouts, these layouts are still designed on the basis of the system designers' intention [13, 14]. Thus, a casual visualization technique is needed for a wide variety of situations.

Several pen-based computers have been developed. At the consumer product level, PDAs and tablet PCs have become common. These computers allow users to interact with the computer environment by means of simple sketch manipulations [4, 5, 6] and are effective for casual and rough inputs such as sketching in the real world. These sketch-based systems allow users to perform 2D drawing manipulation of 3D computer graphics (CG) creations, so the difficulties of 3D CG are avoided. Characteristically, the manipulations required for these systems are simple and similar to drawing a stroke on a piece of paper with a pen. Sketch [4] users can draw 3D curves by performing 2D manipulations. This system calculates a 3D curve by combining a 2D stroke and a shadow stroke. Users of Harold [5] and Tolba [6] can create flat models in a 3D space using sketch-based manipulation, effectively creating a 2.5D scene in a 3D space.

### III. ALLEYOOP DESIGN

One design focus is to enable interactive drawing for information retrieval. Such a system would let users operate the system more freely and easily when engaging in creative activities. Another design focus is to provide both rough and detail searches. This combination makes it possible to treat huge amounts of information. We consider the following features as important to achieving our design goal.

#### A. Simple Creative Interaction

Simple creative interaction was important to realize our concepts because this type of interaction provides flexibility. Integrating drawing with information retrieval enables a user to use the whole workspace interactively, much like drawing a picture on a piece of paper. In such an environment, the user can draw layouts to display information nodes and keywords related to the nodes anywhere within the workspace. While the system treats many types of information (e.g., images, movies, 3D models, documents, and Web pages) as visualized information nodes, we used mainly image data as the information node. An information node means a piece of information on the screen. Also, depending on how a line is drawn, the user can create either simple or complex queries for information retrieval. All the elements related to information retrieval such as layout design and keyword drawing should be realized flexibly.

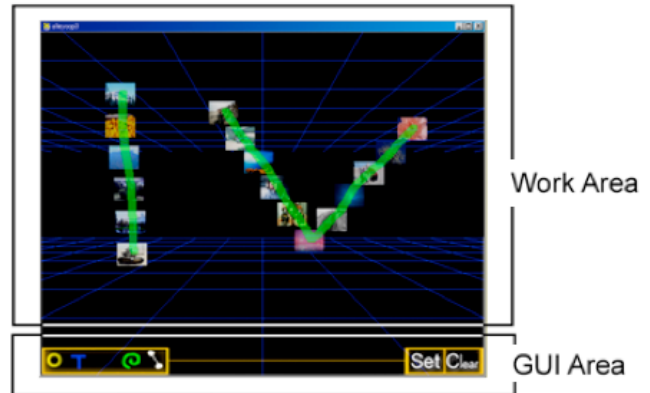


Figure 1. System Overview: The system is divided into two areas, work and GUI areas. Some GUI buttons are included in the GUI area to control pen attributes.

#### B. Rough and Detail Searches

Combining rough and detail search methods is one important element, since we have to treat huge amounts of information. This combination has been supported by conventional Web search engines such as Google and Yahoo. Web users performing searches first collect related information roughly by setting queries and then see each information node in detail one by one. The Alleyoop system also treats huge amounts of information, so it should provide these rough and detail processes. Our system arranges these processes by sketch manipulations (setting an area and creating a layout). In such an environment, the user first creates an outline to collect information nodes roughly and then creates an original layout to see each information node in detail.

#### C. Node Animation

Node animation plays an important role in our information retrieval. For example, through the animation, users can know the relationship between node and keyword dynamically and retrieve extra information. Our system integrates node animation with the user's drawing.

### IV. SYSTEM OVERVIEW

Figure 1 shows the user interface for our system. The system is divided into two areas: work and graphic user interface (GUI) areas. The GUI buttons at the bottom of the figure allow the user to choose pen attributes and draw four types of strokes (area, text, layout, and relation strokes). The first button is for the *area pen*, which is used to draw a closed curve. The second button is for the *text pen*, which is used to draw a keyword. The third button activates the *layout pen*. The user can create a 2D layout by drawing a layout stroke. The fourth button is for the *relation pen*, which allows the user to add relationships between different pieces of a visualized information node.

The interaction is divided into two phases: rough and detail searches. The area and text strokes are used for the former and the layout and relation strokes are used for the latter. The color of the stroke depends on the color of the button, so the user can use stroke colors like paint colors.

The buttons to set and clear GUIs are located in the area. The “Set GUI” button is used to set visualized information alongside the user’s strokes and the “Clear GUI” button is used to clear the work area.

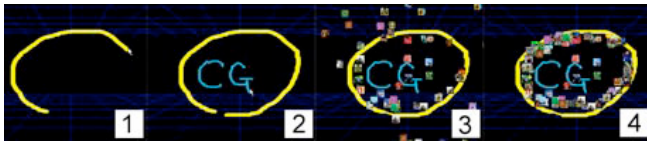


Figure 2. Rough search: A user first draws a closed curve with the area pen and a keyword inside the curve with the text pen. As the system recognizes the area and keyword, related information nodes appear inside the area.



Figure 3. Drawing a Venn diagram: By continuously drawing crossed curves and keywords, a user creates a Venn diagram and can then retrieve information by forming a complex query (1, 2, 3, 4). It is possible to create a Venn diagram between a drawing and an image that contains character information (5, 6).

A. Rough Search

A user can collect information nodes roughly by using the area and text pens. In spite of the simple interactions, it is possible to form complex queries through a series of continuous drawings. The system can also be used in combination with other types of datasets, such as an image dataset.

1) Drawing a closed curve and a keyword

A user starts interacting with the information space by drawing a closed curve with the area pen (displayed as a yellow stroke) and a keyword with the text pen (displayed as a blue stroke). The closed curve provides an area where information nodes related to the keyword will be collected and the keyword provides a query to search for related information nodes from a database. Figure 2 shows an example of simple information retrieval through drawing. First, the user draws a closed curve, and then draws a keyword CG inside the curve (1, 2). The system recognizes the area and keyword, so information nodes related to the keyword move to the closed curve (3, 4). The related information nodes are moved with a force depending on the distance between the node position and the center of the curve. As a result, the information nodes related to the keyword are collected inside the area.

2) Venn Diagrams

By continuing to use simple drawings, a user can form more complex queries. Figure 3 shows an example of

creating a Venn diagram by continuing to draw closed curves and keywords. For example, when a user retrieves information that has two keywords CG and 3D, the user first draws a closed curve and CG (1), and then draws another closed curve and 3D (2). Information nodes related to both keywords appear in the shared area of the Venn diagram (3, 4). From the information distribution results on the Venn diagram, the user can retrieve information roughly and recognize relationships between nodes. Moreover, the user can use an image to create a Venn diagram by combining drawings with image data. In this case, a file name or contained character information becomes a query for the Venn diagram. The figure shows an example of creating a Venn diagram between an image and a drawing (5, 6). In this example, related nodes with both VR and the file name appear.

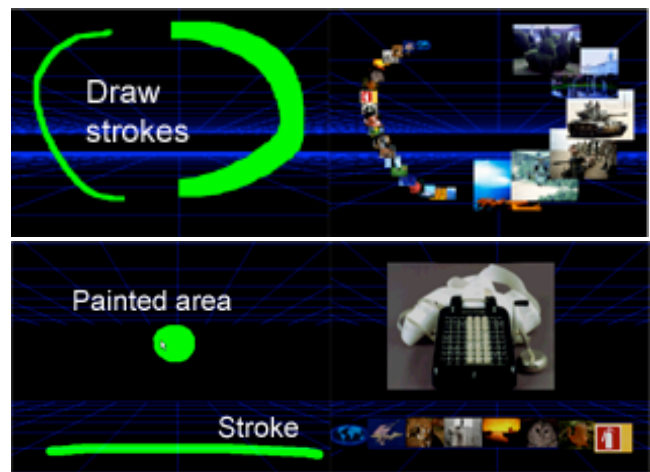


Figure 4. Detail search: The user can create a layout by drawing a layout stroke to see information nodes in detail. It is possible to control the information output by selecting the appropriate pen width (top) and adjusting the painting area (bottom).

B. Detail Search

After collecting related information nodes roughly, a user can see them in detail by using an original layout the user created. The user can create a layout with the layout stroke and add relationships between nodes with the relation stroke.

1) Layout Design

Figure 4 shows how the user can create a 2D layout by selecting the layout pen and drawing a stroke with it (displayed as a green stroke). Since a set of images is automatically displayed alongside the user’s strokes, he/she can create a layout that freely displays the images in the work area. Thus, the user can place information nodes onto drawn words and pictures. In addition, as the pen width is directly related to the image size, larger images appear alongside the stroke if the user draws a bold stroke, while smaller images appear if he/she draws a thin stroke in the work area (Fig. 4 (top)). A painted area is recognized as one big point, so the system positions a large image alongside the painted area (bottom). Thus, the user can select a large image when painting to see it in more detail (Fig. 4 (bottom)).



Figure 5. Relation stroke: The user can create a layered layout by drawing relation strokes (top). Multiple information can be connected to single information nodes at the same time (bottom).

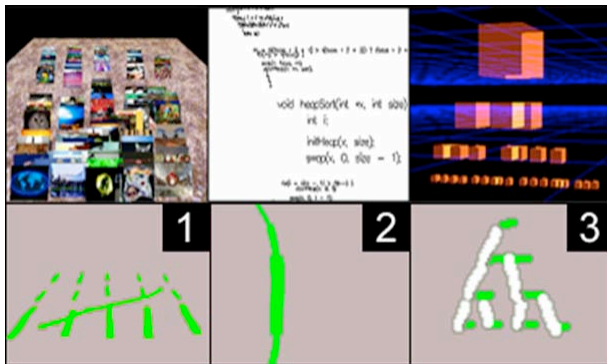


Figure 6. Layout creation with our system. Each layout was created by drawing simple strokes.

Conventional information visualization systems enable users to visually search and browse through a layout to locate grouped information. While these layouts are pre-defined and useful in applications considered by designers, users cannot freely change or redesign them. In contrast, our system integrates information visualization with sketch manipulations and allows users to create a layout freely and easily.

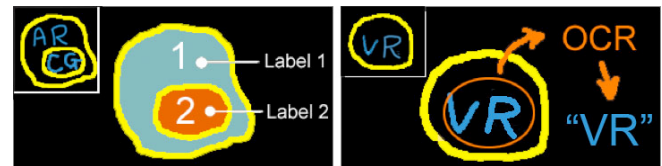
2) Relationships between Nodes

Figure 5 shows how the user can add relationships to connect focused images by drawing a relation stroke (displayed as a white stroke) from one image to another. As the pen width directly corresponds to the relationship value, the user can establish and control relationships between images. The clicked image in the figure becomes the parent for the connected image, and the system establishes the relation through the pen width. This manipulation is especially useful in grouping new information, such as digital photographs and user creations such as painted images or 3D models. Users can now easily take pictures using digital cameras and, as a result, have a huge amount of original image data in their computers. This stroke is useful for such an environment. Moreover, users can create simple information visualization layouts through continuous, straightforward manipulations with our relation techniques (Fig. 6).

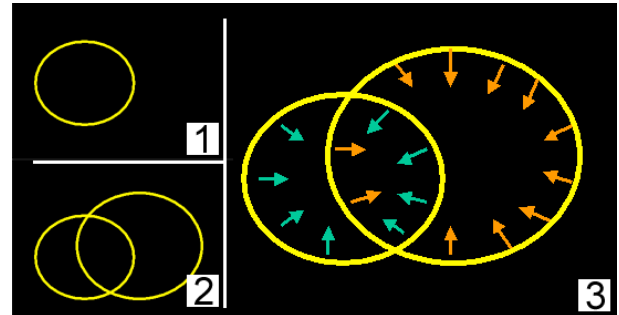
V. IMPLEMENTATION

Next, we consider the implementation of the Alleyoop system. Users' drawing manipulations are reflected in the

bitmap data that make up the workspace. The workspace has four bitmap layers and each layer is used for each pen attribute. The system is basically realized through simple sets of image processing.



(a) The system labels the user's drawing area (left) and then recognizes keywords by using an OCR library (right).



(b) Labeling of a Venn diagram. The system uses mouse trace data to determine the shared area and its keyword.

Figure 7. Implementation of rough searching. The implementation is divided into two parts: labeling of the user's drawing area and recognition of a keyword.

A. Rough Search

When the user draws an area with the area pen and a keyword with the text pen, the system stores four types of data (stroke-ID, mouse-trace, area-bitmap, and keyword-bitmap data).

When the user pushes the "Set GUI" button after drawing a closed curve and a keyword, the system starts the calculations for the user's drawing. For the closed curve area, the system labels the inside of the area (Fig. 7 (a-left)). This labeling process depends on the mouse trace data, which the system uses to determine the inside or outside area. For the keyword bitmap data, the system sends the data to an optical character recognition (OCR) library and recognizes the keyword meaning (Fig. 7 (a-right)). When the user creates a Venn diagram through continuous drawing of a closed curve and a keyword, the system uses the mouse trace data to determine the shared area and its keyword (Fig. 7 (b)). In the event the combination includes an image, the system labels the inside of the image area with the same label as the drawn curve. The image is thus treated as the drawn curve. After the calculations, the system contains three types of data (area ID, area position, and keyword) for each labeled area.

Next, calculations for the node animations start. Each node is subject to a force that depends on the node's relationship with the keyword. This calculation is done by a spring model [9]. The node then moves until it is in the appropriate area for the relationship. The system has an original database containing three types of data: node names, keywords, and relation levels. The database is loaded when the system starts.



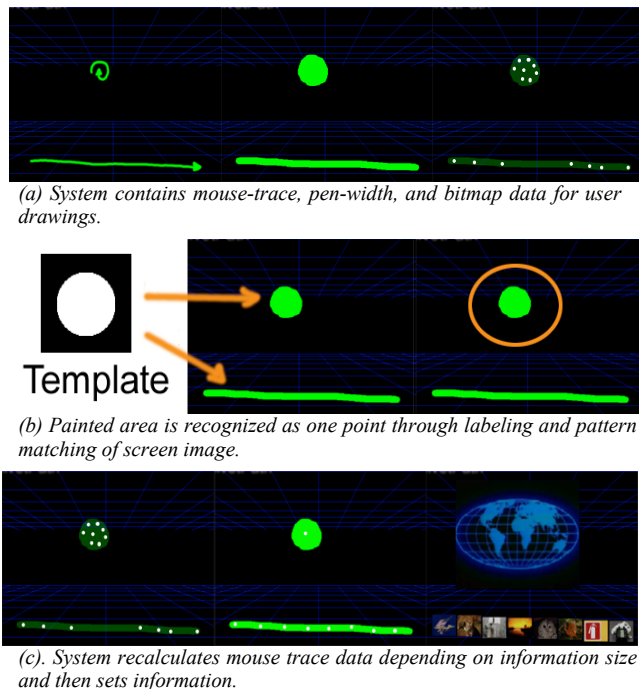


Figure 8. Detail search implementation and drawing area calculation.

### B. Detail Search

As Fig. 8 shows, for a detail search the system stores three types of data (i.e., mouse-trace, pen-width, and bitmap data from the screen) for a set of layout strokes (a). The bitmap data is labeled and then matched to simple template data (b). The system recognizes what shape was drawn (e.g., line or circle) through template matching with the labeling data. If the results for labeling are different from the template pattern, the system recognizes strokes as lines and sets visualized information. Next, the trace data is recalculated to maintain the same distance according to pen width, enabling a painted circle to be recognized as one big point (c). The amount of visualized information shown on the screen depends on the stroke length. Less information is visualized if the total length of the line on which visualized information is laid out is longer than the stroke. As a result, visualized information is automatically set along with the user's strokes. These calculations to set information are done when the user draws strokes and pushes the "Set GUI" button.

In connecting with a relation stroke, the system recognizes visualized information under the starting point as a parent and visualized information connected with the relation stroke as a child of the scene graph. The system defines the size and position parameters of the child by taking the stroke width into account. The child image that is included in the layout becomes half the size of the parent, because they are connected with a double-sized pen. These results are saved by the system database.

## VI. DISCUSSION

Here, we discuss user interactions with Alleyoop based on comments made by visitors to our demonstrations.

### A. Rough Search

In our demonstrations, visitors quickly understood the system concepts and interaction methods. It was not difficult for them to recognize the relationships between information nodes and keywords directly through Venn diagrams, or to use the whole workspace for free drawing. Most visitors were able to create simple Venn diagrams and set related nodes onto them after a simple demonstration. While conventional information retrieval systems require users to input a keyword to a dialog box field, our system allows them to use the entire workspace and retrieve information through creative activities. Since the system facilitates creative activities, we expect its users will be able to create more original and effective drawings for information retrieval. We also received good reactions from visitors regarding the combination of drawings and images to create a Venn diagram. Our system can use various types of information included in images, such as characters and words. Through sketch manipulations, a user can retrieve information without drawing a keyword. Moreover, the user can use other types of data, such as real-world information, to create Venn diagrams. If the information contains text and characters, these information elements are used as queries for information retrieval. Figure 9 shows an example of using real-world information as a Venn diagram element. The user first captures real-world information through a digital camera attached to his/her computer (1, 2) and then draws a closed curve around a keyword on the captured data and another closed curve to collect information nodes (3, 4). As the system recognizes the keyword inside the first closed curve, related information nodes appear inside the second curve.

### B. Detail Search

Most visitors designed original layouts through simple manipulations. Although none of them created a complex layout, they were able to easily create simple layouts and created words and pictures to visualize grouped information. This indicated that they quickly understood our system concepts and interaction methods. Our system was designed to enable users to design simple layouts and freely redesign layouts. The visitors were able to casually create simple or unusual layouts not requiring special design skills and encountered no difficulties in this casual use. Allowing users to design and create layouts freely in and of itself, however, does not ensure an easy design process. Accordingly, our system provides sketch and paint techniques instead of complex parameter settings. This enables even a simple stroke to become a kind of layout and thus allows design difficulties to be avoided. We also showed through simple demonstrations that both sketch and paint interactions could be used to directly control parameters (e.g., scale and relation). The visitors reacted particularly favorably to the relation stroke because it made it possible to handle the user's original data set while controlling the relation rate by adjusting the pen width. They generally held their original data set in a folder named "Related Data", and all were able to establish relationships between data and create simple databases using only the relation strokes.



Figure 9. Combination with real-world information: Capturing real-world information as a picture (1, 2) and drawing two closed curves (3, 4). One curve (shown in yellow) means an area and the other (shown in blue) means a keyword.

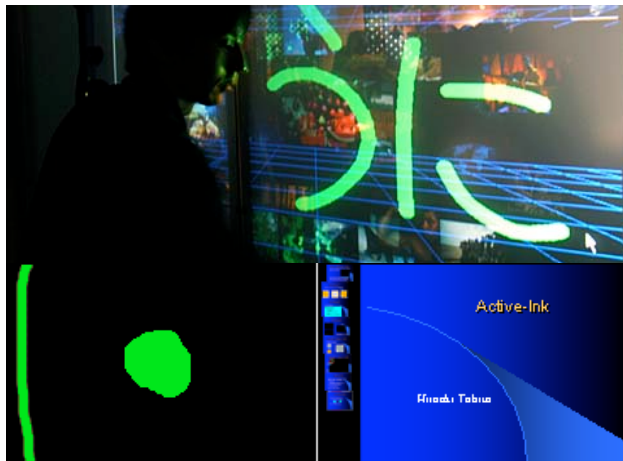


Figure 10. Combination with wall-type display. A user sets data by drawing a stroke on a wall-type display.

Some visitors had questions related to the system's lack of effectiveness with detailed manipulations. A few of them also said they would have preferred more complex layouts and relationships. However, our goal was to enable simple and convenient interactions that would allow casual users to freely utilize the system. As complex manipulations and GUIs would have increased the complexity of the layout and relations, we did not include these in our system.

### C. Communication

It is also possible to combine our system with AR systems based on tablet computers [7] and with ubiquitous computing environments. As the user generally interacts with the computer to display and change visualized information in these environments, information layout techniques are as important as interaction techniques. To enable more natural interactions, the visualized information size and position need to change depending on the situation and type of computer to allow more natural interaction with the information space through both visualization techniques and AR systems. Also, as the interactions should be simpler than for normal computing in these systems, a combination of our system and AR systems promises to be quite useful. For example, in case of using wall-type AR systems, the user can use the wide information space freely (Fig. 10 (top)). Our system also provides good support for displaying information for presentation in meetings (Fig. 10 (bottom)). Here, the user can draw the most effective layout according to the situation and type of information.

### D. Future Work

Our system is so far only at the prototype stage, so we are planning to further improve it by combining with other types of database. Then, we will conduct practical user tests.

As a subject for future work we plan to combine our system with AR system to enable it to support more natural interactions [10]. Combining it with a gesture recognition system would be a particularly effective way to achieve this.

## VII. CONCLUSIONS

The Alleyoop system enables users to retrieve information through sketch manipulations. The system is divided into two phases: rough and detail searches. We have described our design concepts and a prototype of the Alleyoop system. Although the interactions based on sketch manipulations are very simple, users can use them to set complex queries and create original layouts freely.

## REFERENCES

- [1] H. Tobita. Catenaccio: interactive information retrieval system through drawing. In Proceedings of AVI2006. pp.79-82 .
- [2] H. Tobita. VelvetPath: Layout Design System with Sketch and Paint Manipulations, In Proceedings of EUROGRAPHICS2003 Short Presentations, pp. 137-144, 2003.
- [3] S. K. Card, J. D. Mackinlay, and B. Shneiderman. Readings in Information Visualization : Using Vision to Think, 1999.
- [4] R.C. Zeleznik, K.P. Herndon, and J.F. Hughes. An Interface for Sketching 3D Curves. *SIGGRAPH '96 Proceedings*, pp. 163-170, 1996.
- [5] J.M. Cohen, J.F. Hughes, and R.C. Zeleznik. Harold: A World Made of Drawings. *NPAR2000 (Symposium on Non-Photorealistic Animation and Rendering)*, pp. 83-90, 2000.
- [6] O. Tolba, J. Doresey, and L. McMillan. Sketching with Projective 2D Strokes. *Proceedings of UIST '99*, pp. 149-157, 1999.
- [7] J. Rekimoto. Pick-and-Drop: A Direct Manipulation Technique for Multiple Computer Environments. *Proceedings of UIST'97*, pp. 31-39, 1997.
- [8] G. G. Robertson, J. D. Mackinlay and S. K. Card. Cone Trees: Animated 3D Visualization of hierarchical information. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '91), pp. 189-194, 1991.
- [9] R. Davidson and D. Harel. Drawing Graphics Nicely Using Simulated Annealing. *ACM Transactions on Graphics*, Vol. 15, No. 4, pp. 301-331, 1996.
- [10] Rekimoto, J. and Ayatsuka, Y. Cybercode: Designing Augmented Reality Environments with Visual Tags. In *Proceedings of DARE 2000*, pp. 1-10, April 2000.
- [11] G. W. Furnas. Generalized fisheye views. Proceedings of the ACM Tran. on Computer-Human Interaction, Vol. 1, No. 2, pp. 126-160, 1994.
- [12] B. B. Bederson, J. D. Hollan, K. Perlin, J. Meyer, D. Bacon, and G. Furnas. Pad++: A Zoomable Graphical Sketchpad for Exploring Alternate Interface Physics. *Journal of Visual Languages and Computing*, Vol. 7, No. 1, pp. 3-31, 1996.
- [13] L. Zhicheng, J. Stasko, T. Sullivan. SellTrend: Inter-Attribute Visual Analysis of Temporal Transaction Data. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1025-1032, 2009.
- [14] B. B. Bederson, B. Shneiderman, and M. Wattenberg. Orderedandquan- tum treemaps: Making effective use of 2D space to display hierarchies. *ACM Transactions on Graphics*, 21(4), pp. 833-854, 2002.

# Interactive Hand Gesture-based Assembly for Augmented Reality Applications

Rafael Radkowski

Heinz Nixdorf Institute, Product Engineering  
University of Paderborn  
Paderborn, Germany  
rafael.radkowski@hni.uni-paderborn.de

Christian Stritzke

University of Paderborn  
Paderborn, Germany  
cstritzk@uni-paderborn.de

**Abstract**—This paper presents an Augmented Reality (AR) assembly system for the interactive assembly of 3D models of technical systems. We use a hand tracking and hand gesture recognition system to detect the interaction of the user. The Microsoft Kinect video camera is the basis. The Kinect observes both hands of a user and the interactions. Thus, a user can select, manipulate, and assemble 3D models of mechanical systems. The paper presents the AR system and the interaction techniques we utilize for the virtual assembly. The interaction techniques have been tested by a group of users. The test results are explained and show that the interaction techniques facilitate an intuitive assembly.

**Keywords** - Augmented Reality; Interaction; Interactive Assembly

## I. INTRODUCTION

The Augmented Reality (AR) technology is a kind of human-computer interaction that superimposes the visual perception of a user with computer-generated information (e.g., 3D models, annotations, and texts). AR presents this information in a context-sensitive way to the user. 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. In the field of mechanical engineering, AR applications are explored for assembly, service, maintenance, and design reviews [1] [2].

Intuitive interaction techniques are a major aspect of AR applications. Intuitive means, a user is able to manipulate virtual objects without the aware utilization of prior knowledge. For that purpose, different interaction techniques and concepts have emerged in the field of AR.

A major driver for intuitive interaction is vision-based hand gesture recognition. Computer vision (CV) algorithms analyze a video stream, detect the user's hands, and determine a gesture. The gesture fires an interaction function of the AR application that manipulates (e.g., translate, etc.) a virtual object.

Hand gestures and computer vision-based hand gesture recognition are assumed to approximate the natural interaction of humans closely [3]. A user can interact with a virtual object like s/he interacts with a physical object using his/her hands. One advantage is, that a user does not need to wear or carry any technical device in his/her hand.

However, hand gestures and hand gestures recognition is still a challenging research field. In the field of AR, the techniques have not left the research laboratories until today.

One reason is the need of technical devices that are attached to the user's hand in order to track it. The user also still act as operator of a machine; the interaction is not intuitive [4].

New types of video cameras like Microsoft Kinect facilitate the free-hand interaction. This video camera, particularly its images are utilized to detect the hands of a user and the hand gesture. However, it is necessary to explore interaction techniques and interaction metaphors for the interaction with virtual objects.

This paper presents an AR assembly system that facilitates the virtual assembly of virtual parts. A Kinect video camera observes the user and its interactions; the position of the hands and their gestures are recognized. No addition devices need to be attached to the hands of the user. This paper explains the system and describes the interaction metaphors, which are suitable for a free-hand interaction. The metaphors have been tested by a group of users. The results are presented.

This paper is structured as following. In the next section, the relevant related work is reviewed. Afterwards the AR assembly system is introduced as well as the interaction metaphors for selection, manipulation, and the virtual assembly of 3D models. Section 4 describes the user test. The paper closes with a summary and an outlook.

## II. RELEATED WORK

The related work addresses the field of hand gesture-based interaction in AR applications.

One aim of many AR applications is to provide a natural and intuitive interaction interface to the user [3] [5] [6]. Like physical objects, a user should be able to grasp a virtual object like a real object. The user should grasp an object with his/her fingers, pick it up, and place it at every desired location. Therefore, different approaches exist.

One of the first systems is introduced by Buchmann et al. [3]. Their system, called FingARtips, is a gesture-based system for the direct manipulation of virtual objects. They attached fiducial markers on each finger to track the fingertips and to recognize the hand gesture. This solution allows to pick up virtual objects. However, it uses markers.

Reifinger et al. introduce a similar system [5]. The authors utilize an infrared tracking system for hand and gesture recognition. Infrared markers are attached to the fingertips and hands of a user. A tracking system detects the markers and a computer-internal hand model is build using this data. Thus, the user is able to grasp virtual objects like real objects.

Lee et al. introduces a system that does not utilize physical markers [6]. The authors use a feature-based tracking system. Computer vision algorithms identify hand features. Thus, the system detects and tracks the user's hand. This allows a user to attach a 3D model virtually to his/her hand, to move it, and to place it on different positions. However, a realistic grasping action is not been realized.

Siegl et al. introduces the concept of 3D-cursors as interaction metaphor in AR applications [7]. The user of an AR application is able to indicate a point in space with his/her hand. A vision-based system recognizes the hand. This way, the indicated position is calculated.

We introduce our own hand recognition system in [8]. In this previous work, we investigate how important the visibility of the user's hand for interaction purposes is. Furthermore, the hand recognition system is introduced. Further systems can be found in [9], [10], [11], and [12]. These examples provide only an overview about the research field; it would exceed the size of this work to present all systems. However, two findings can be stated: first, intuitive, natural interaction is one aim of AR. However, the working systems utilized physical markers. Second, there are many efforts to realize a computer-vision based hand gesture recognition system that works without any physical markers. A practical solution does not exist until today.

### III. AUGMENTED REALITY ASSEMBLY SYSTEM

This section describes the AR assembly system for the virtual assembly of virtual parts of mechanical systems. The provided interaction techniques facilitate to manipulate and to assemble virtual parts. All interactions are carried out by hand movements and hand gestures, without any devices attached to the user's hand. The section starts with an overview of the AR system, which includes a presentation of the used hardware and software. Afterwards the interaction techniques for selection and manipulation tasks are introduced as well as the interaction for the virtual assembly.

#### A. Overview

Figure 1 shows an overview of the hardware setup of the AR application. A table is the main working area. The user stands on a fixed position in front of the table. We use a monitor-based AR-system. It consists of a 24" widescreen monitor and a video camera. The video camera captures images of the scene in front of the user. It is located next to the user, close to the user's head. It simulates a camera attached on a head mounted display. We have decided for a statically arranged camera in order to get comparable test conditions. It is a Creative Live Cam Video IM Ultra webcam (1280 x 960 pixel at 30 fps). The user observes the augmented scene on the screen of the monitor. For tracking the ARToolKit is used, a pattern-based tracking system [13]. Altogether, the setup represents a common monitor-based AR application.

The Kinect video camera stands opposite to the user. It observes the user and the user's interactions. The camera is aligned into the direction of the user. It captures RGB color images with a resolution of 640 x 480 pixel and 12bit depth images. The user does not see these images during s/he uses

the AR application. The working area of the Kinect camera is arranged manually to the working area of the user. Therefore, the user put his hand to two corners of a control image. A region of interest is specified with respect to these corners. Thus, the camera as well as the user has to stand on a fixed position after alignment.

As computer we used a PC with an Intel Xeon processor, 3.5 GHz, 6GB RAM, and a NVIDIA Quadro 5000 graphics processor.

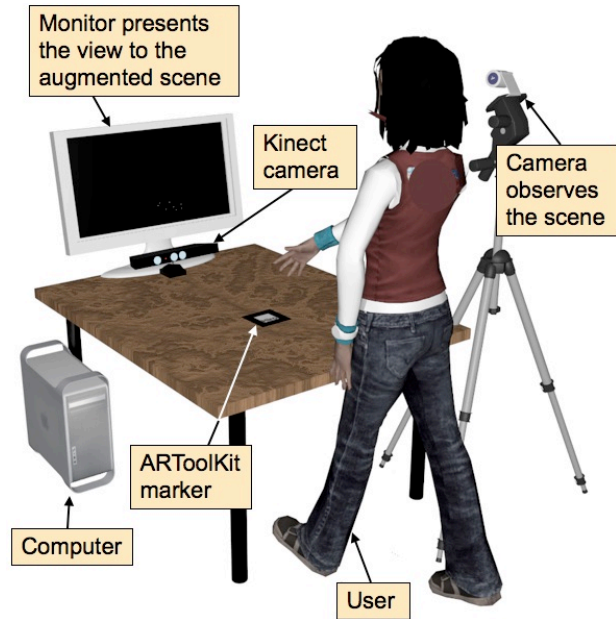


Figure 1. Overview of the hardware setup of the AR system

Two software tools are used to run the AR assembly system: a hand gesture recognition software and an AR application. The hand gesture recognition software detects the hand position in space and the hand gesture. It detects five gestures: fist, open hand, closed hand, index finger, and a waving gesture. The software is based on OpenCV (<http://opencv.willowgarage.com/>), an open source computer vision library.

The hand position and gesture are submitted as messages via UDP/IP to the second tool, the AR application. The AR application is based on OpenSceneGraph, an open source scene graph library ([www.openscenegraph.org](http://www.openscenegraph.org)). It facilitates the rendering of 3D models, provides functions for collision detection and supports interaction. For our purpose, the AR application provides the following functions: selection of 3D models, translation, rotation, scale, change of attributes, and virtual assembly. The gesture recognition software and the AR application have been described in detail in [14].

#### B. Interactive Selection

Figure 2 shows the view of the user that is presented on the screen. It displays the main view of the application. The main view shows the virtual parts, which need to be assembled. In addition, multiple virtual button icons are shown. The button icons allow selecting a distinct function (i.e., translation, rotation, etc.).

The interactive selection allows a user to select a 3D model or a manipulation function. The main interaction object is a virtual cursor; a 3D sphere (3D cursor). The 3D sphere indicates the position of the user's hand. It follows the movement of one hand in three dimensions. To select a 3D model the user has to move his/her hand, in particular the 3D sphere to a 3D model. The selection is implemented as collision detection between the sphere and a virtual part. This collision is considered as a selection when the user applies a fist gesture. When the hand is opened, the model is released. Thus, grasping is simulated. As visual clue, a selected 3D model is colored yellow, a selected menu item is colored red.

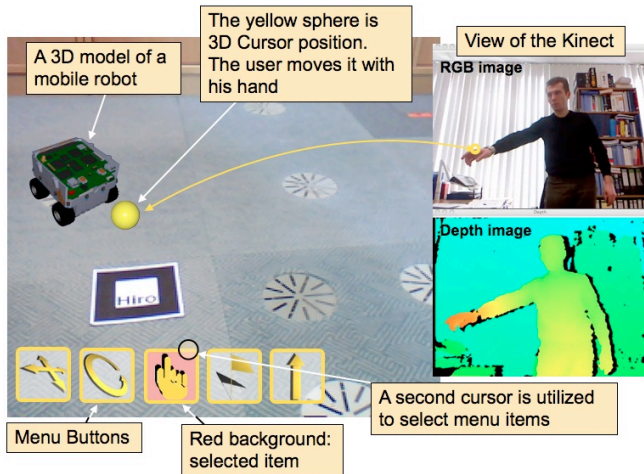


Figure 2. Main view of the AR application.

In addition to the 3D sphere, a 2D circle is used as selection object (also called 2D cursor). It is used to select a button icon in the menu. Usually, the 2D cursor is invisible. It appears only when the user moves his/her hand close to the button icons. Therefore, a region of interest (ROI) is specified that encloses the menu. If the hand of the user and the 3D sphere touches this ROI in screen coordinates the 2D cursor appears. To activate a function, the user has to move the circle above a button and has to wait for two seconds.

### C. Interactive Manipulation

Interactive manipulation includes the translation, rotation, scaling, and the change of attributes (e.g. color) of a 3D model. The function can be applied to each 3D model after selection.

Two modes of operation exist: a so-called direct mode and a precise mode. The user can switch between both modes.

#### 1) Direct mode

The direct mode allows to move and to rotate an object directly. The user can grasp it virtually and move it. Therefore, the user selects a 3D model. This 3D model is being attached to his/her hand and follows all movements. Furthermore, if the user rotates his/her hand, the model follows this rotation, limited to one degree of freedom.

The direct mode facilitates a fast movement of 3D models in the working area. In addition, this technique

appears to be intuitive, because it meets a common grasping / pick & place operation. Unfortunately, it does not allow a precise alignment of virtual objects. Scaling and the change of attributes are also not supported in this mode.

#### 2) Precise Mode

The precise mode facilitates an accurate translation, rotation, and scaling of virtual parts. 3D models are utilized as visual clues and interaction objects in order to support these functions.

**Translation:** Figure 3 shows the translation of a 3D model. The figure shows the view of the user. After selecting the function, a 3D model of a coordinate system appears above the virtual object that should be moved. This coordinate system indicates the possible moving directions.

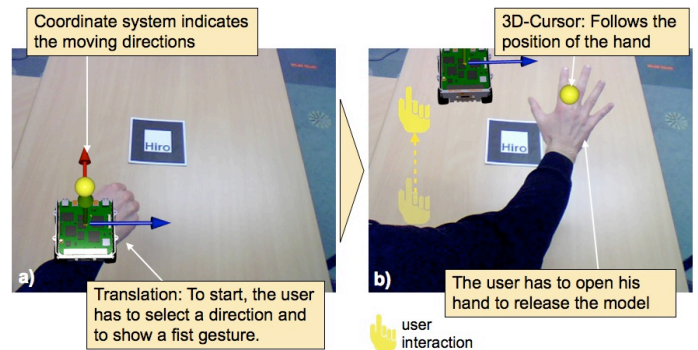


Figure 3. a) To translate a 3D model the user has to grasp an axis of the coordinate system that appears above the 3D model, b) The user has to open his hand in order to release the 3D model.

The user has to choose one axis of the coordinate system (collision detection) using the yellow 3D cursor. To translate the model, a fist gesture has to be performed. This should simulate grasping. This starts the translation into the desired direction. To move the 3D model the user has to move his/her hand into the desired direction. The movement of the hand is assigned to the translation of the model. The user can select a scale factor that slows down or speeds up the movement. Usually, a scale factor between 0.5 and 1.5 is selected.

**Rotation:** After selecting this function and a 3D model, a virtual coordinate system appears above the 3D model. To rotate the object, the user has to grasp one axis of this coordinate system. Therefore, s/he uses the 3D cursor. To select the axis a fist gesture need to be performed. Doing this, the rotation starts. To rotate the 3D model the user rotates his/her hand about the selected axis. Every movement of the hand is transformed into a rotation. The start point is the angle, at which the fist gesture has been shown. It works similar for the rotation about the other axis. The interaction technique should simulate a rotation of an object using a lever. The lever is utilized to slow down or to magnify the rotation.

**Scaling:** The scaling works similar to the rotation and translation. After selecting this function and a 3D model, a virtual box and a coordinate system appears above the 3D model (Figure 4). To coordinate system indicates the

possible scaling directions. The box is a visual clue that helps to recognize the scaling factor. To scale the 3D model the user has to select an axis using the virtual 3D cursor and to perform a fist gesture. Then s/he has to move his/her hand along this axis. The movement is multiplied as scaling factor to the 3D model. In addition to the 3D model, the box is also scaled. If the 3D models are unshaped, the box facilitates to recognize the scaling factor.

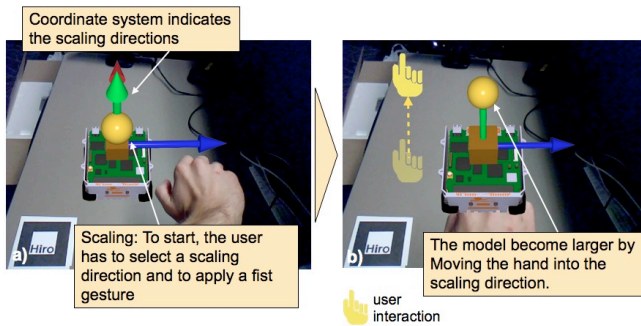


Figure 4. Scaling of 3D objects

D. Interactive Virtual Assembly

The main function of the virtual assembly system is to provide interactions for the assembly of virtual parts. Assembly means in this case that two or more models join together when they meet on a specified position. The assembly system utilizes the selection and manipulation functions presented before. In addition to these functions, a pre-defined mode switch simplifies the assembly of two models. In the following, the assembly is explained using an example of an axle and a ball bearing. It works also for all virtual parts in the same way.

The virtual assembly is based on a so-called port concept [15]. A port is a distinct position on the surface of a 3D model that is annotated by a joint. A joint limited the degrees of freedom between two virtual parts. An octahedron on the surface of the 3D model visualizes this joint (Figure 5). Five types of joints are implemented. Each type limits a different degree of freedom: hinge joint, ball & socket joint, linear bearing, rotation bearing, and a fixed joint. The different types are visualized by different colors of the octahedron. The joints and its position are specified in a pre-process.

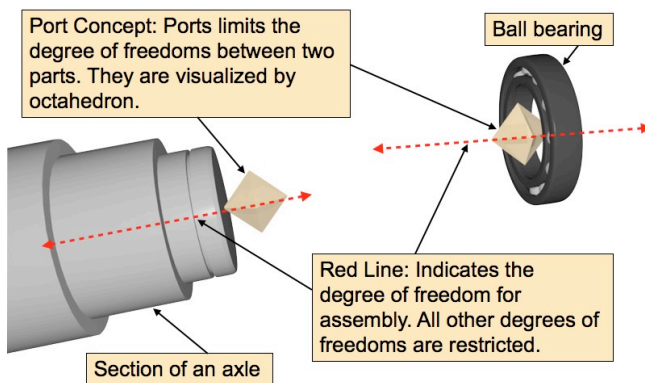


Figure 5. A port concept is used to assemble two virtual parts.

To assemble two parts the user has to move closely the two octahedrons of two parts. Two parts that using the same type of joint can be assembled only.

For the assembly task itself a two-step interaction is used [16]. Figure 6 shows the first step. The figure shows an axle and a ball bearing. Task of the user is to assemble the ball bearing wheel on an intended section of the axle.

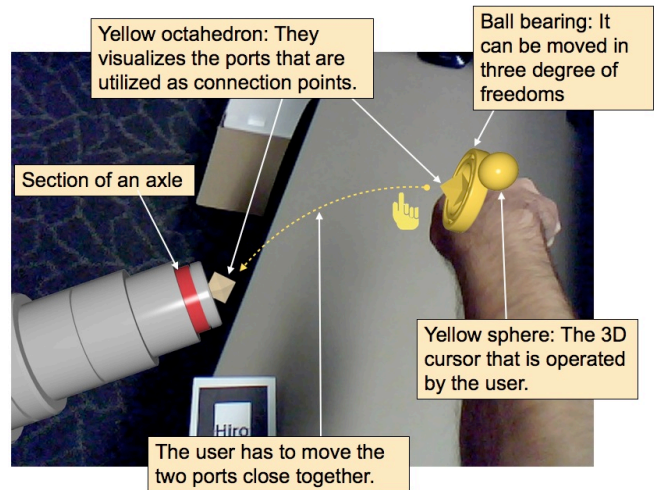


Figure 6. The assembly is carried out in two steps. In the first step, the user can move the 3D models in three degree of freedoms

In the first step, the user has to move both parts close together. Therefore, the mentioned direct mode can be used. Bounding boxes surrounds the octahedron. As soon the bounding boxes collide a mode switch is applied automatically. By this, all degrees of freedom of the two parts, which become fix after assembly, are being aligned. In this mode the user can move the part along the remaining degree of freedom only. In this example (Figure 7): as soon the ball bearing is close to the axle, the ball bearing is moved to the center axis of the axle. Furthermore, the ball bearing can be only moved along the center axis of the axle.

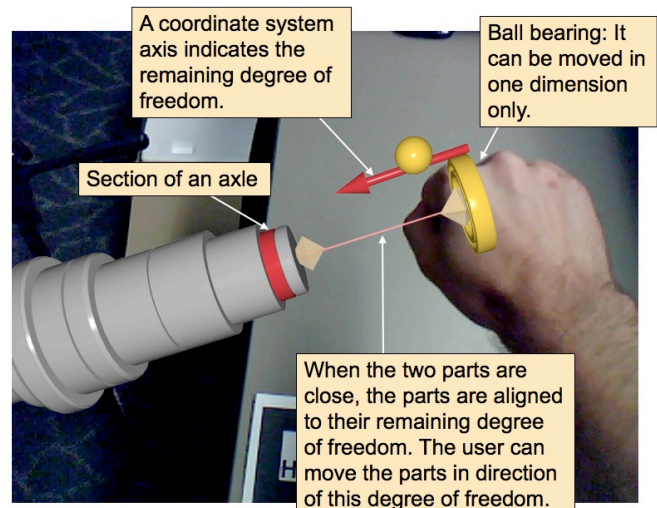


Figure 7. Second step of the virtual assembly. The user can only move the part in the remaining degree of freedom.

After the part is on the desired position, the user can release the part. Thus the ball bearing remains assembled on the axle and both parts can be moved as one group.

IV. USER TEST

The AR assembly system has been tested by a group of users. The aim of the test was to explore whether the interaction techniques facilitate the manipulation and the assembly of virtual parts. Furthermore, we wanted to discover whether the entire system meets the expectation of the users. We do not analysis the quality of the assembly, which would be necessary when using this AR application as simulation tool for real assemblies. In the following, the process is described and the results are presented. The section closes with a discussion of the results.

A. Process

During the test, the users should carry out an assembly task. Task was to assemble gear wheels, ball bearings, and clogging on an axle. The assembly tasks have been introduced to the test users by images. Each image has shown one assembly step (one part = one step). In summary, six parts have been needed to be assembled on the axle.

Before the test has started, the different interaction techniques were presented. Each user has gotten several minutes to practice the interaction techniques. During the test the user could decide on their own, which interaction techniques they want to use and what series of interactions are necessary to assemble two parts.

The test users were 15 students of the departments of mechanical engineering and computer science. No user has experience with hand gesture-based interaction techniques.

We have measured the time a user needed to assemble all parts. In addition, we have used a questionnaire to retrieve the opinion of the users. We asked eight questions (table 1). A Likert scale was used to rate the questions. The scale ranges from 1=“the statement meets my opinion” to 5=“I disagree with this statement” (The questionnaire and the answers were in German).

B. Results

Figure 8 presents the results of the time measurement with respect to the assembly task.

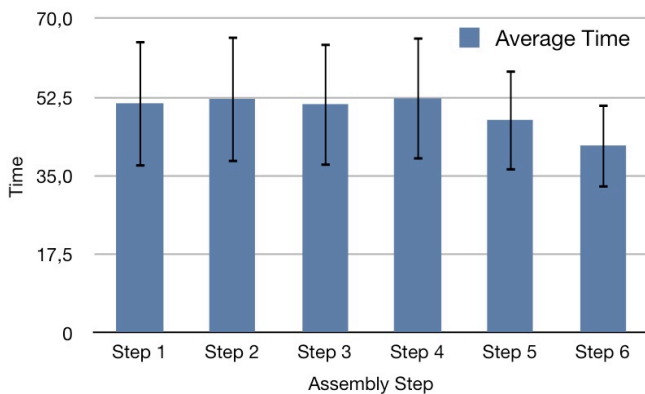


Figure 8. Average time per assembly time

The abscissa displays the six assembly steps, the ordinate displays the time. The bars indicate the average duration time for each particular step. The time measurement for each step has started automatically when two parts in the previous step were assembled. For the first step, the measurement was started manually. It can be observed that there are no significant changes between the different steps. However, there is a large variance.

Figure 9 presents the time measurement with respect to the interaction techniques. The abscissa shows the different interaction techniques, the ordinate the time. The bars show the average time each interaction technique has been in action. The time measurement has started when a user calls the function, it stops when a user has exit a function. The numerical values on each bar indicate the number each function has been called.

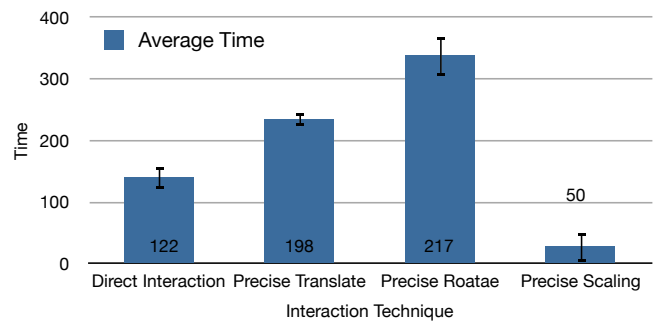


Figure 9. Average time per interaction technique

The results show that the precise interaction techniques demand more time than the direct manipulation techniques. The variance is marginal. More interesting is the number of function calls. It can be observed that the precise techniques have been called more often than the direct techniques.

Table 1 shows the results of the questionnaire. The first column contains the questions. The second column shows the average values of the answer and the third column the variance.

TABLE 1: RESULTS OF THE QUESTIONNAIRE

Question	Av. Result	Variance
1. I can use the manipulation interactions without any problems	1,53	0,52
2. It was easy to select the different 3D mdels	2,27	0,80
3. I understand the mode switch during the assembly task and was able to assemble the 3D models.	2,27	0,59
4. I used the direct mode for large movements and the precise mode to align objects	3,47	1,36
5. Overall, I had no problems to interact with 3D models.	1,93	0,46

C. Discussion

In general, the results of the user test prove that the selected interaction techniques facilitate the virtual assembly of virtual parts. All users were able to assemble the parts in six steps using the provided interaction techniques. All steps

could be completed. The duration of every step is nearly similar (Figure 8). The techniques appear to be controllable from the first time they have been used. A learning phase seems not to be necessary. Only at step 5 and step 6 a slow increase of the assembly time is recognizable. This likely indicates a learning effect, but it is not significant.

Figure 9 shows the average time for an interaction. The results show that the direct interaction is the fastest technique to translate a part. By putting time on the same level as difficulty, direct interaction is the simplest technique. The most difficult technique is the rotation. One reason is that the users need to grab the part several times. During the rotation, they lost the orientation and need to start over. The high number of rotation operations underpins this. The average time of precise scaling is low due to the fact that scaling was necessary in two assembly steps only.

However, there are some drawbacks. We assumed that the user uses the direct mode to move a part onto the workplace, close to the axle. Then s/he should use the precise mode to move the parts to a close distance before the mode switch and the two parts are aligned. The results show that a few users do not understand the intended way. The answer to question 3 of the questionnaire proves that a few users do not recognize this. In addition, the direct interaction was used 122 times. The low number results due to the fact, that six users do not use this technique (Figure 9).

In addition, the users have moved their hands very slow. They operate very carefully during the assembly task.

## V. RESUMEE AND OUTLOOK

This paper presents a set of interaction techniques for the virtual assembly of virtual parts using a hand gesture-based interaction technique. We introduced a set of interaction techniques that allows interacting with virtual parts without using a graspable device. Therefore, we distinguish a direct mode and a precise mode. The direct mode allows fast translation. The precise mode facilitates a precise placing of virtual parts. In our opinion these two modes are necessary to facilitate an interaction with non-graspable parts. The user test gives us a strong indication that the techniques are capable to carry out a virtual assembly task. Finally the work shows that these separate techniques are a good choice for this kind of task.

The future work has two objectives. First, we will carry out an assembly using virtual and physical parts. Until now, virtual parts have been only used. This justifies no AR application. In the next step the users should be able to assemble virtual parts on physical parts.

Furthermore, we will test the precision of the Kinect and the entire AR system. Therefore, we will carry out an experiment with pick & place operations.

## REFERENCES

- [1] Radkowski, R., "What makes an Augmented Reality Design Review Successful?" in Horváth, I.; Mandorli, F.; Rusák, Z. (Eds.): Proceedings of the TMCE 2010 - Tools and Methods for Competitive Engineering, Ancona, Italy, April 12-16, 2010, pp. 499-510
- [2] Azuma, R., "A Survey of Augmented Reality" In: In Presence: Teleoperators and Virtual Environments 6, 4 (August 1997), 1997, pp. 355-385
- [3] Buchmann, V., Violich, S., Billingham, M., and Cockburn, A., "FingARtips: gesture based direct manipulation in Augmented Reality," in Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia (GRAPHITE '04), 2004, pp. 212-221
- [4] Malassiotis, S. and Strintzis, M.G., "Real-time hand posture recognition using range data," in Image and Vision Computing, 26(7), 2008, pp. 1027-1037
- [5] Reifinger, S., Wallhoff, F., Ablassemeier, M., Poitschke, T., and Rigoll, G., "Static and Dynamic Hand-Gesture Recognition for Augmented Reality Applications," in Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments. Springer Verlag, Berlin / Heidelberg, 2007, pp. 728-737
- [6] Lee, T. and Höllerer, T., "Hybrid Feature Tracking and User Interaction for Markerless Augmented Reality," in Proceedings of IEEE Virtual Reality 2008, 8-12 March, Reno, Nevada, USA, 2008, pp. 145-152
- [7] Siegl, H., Schweighofer, G., and Pinz, A., "An AR Human Computer Interface for Object Localization in a Cognitive Vision Framework". In: Int. Workshop on Computer Vision in Human-Computer Interaction (ECCV), Springer 3058, 2004, pp. 176-186
- [8] Radkowski, R. and Wassmann, H., "Using Computer Vision for Utilizing the Human Hand in An Augmented Reality Application". In: Proc. of IADIS Computer Graphics & Visualization, Lissabon, Portugal, 2007, pp. 127-131
- [9] Hackenberg, G., McCall, R., and Broll, W., „Lightweight Palm and Finger Tracking for Real-Time 3D Gesture Control.“ in IEEE Virtual Reality 2011, pp. 19-26, Singapore, 2011
- [10] Lu, Y. and Smith, S., "GPU-based Real-time Occlusion in a CAVE-based Augmented Reality Environment," in 2007 ASME International Design Engineering Technical Conference & Computers and Information in Engineering Conference (CIE), September 4 – 7, 2007 Las Vegas, Nevada, 2007, pp. 1131-1141
- [11] Argyros, A.A. and Lourakis, M.I.A., "Tracking skin-colored objects in real-time," In: Cutting Edge Robotics, 2005, pp. 77-90
- [12] Sung, K. K., Mi, Y., N., and Phill, K.R., "Color based hand and finger detection technology for user interaction," in International Conference on Convergence and Hybrid Information Technology, 28-29 August 2008, pp. 229-236
- [13] Kato, H. and Billingham, M., "Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System," in Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99), San Francisco, USA, 1999, pp. 85-94
- [14] Radkowski, R. and Stritzke, C., „Comparison between 2D and 3D Hand Gesture Interaction for Augmented Reality Applications,“ in Proceedings of the ASME 2011 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2010 Aug. 28-31, 2011, Washington, DC, USA, pp. 1-11
- [15] Shen, Q., „A Method for Composing Virtual Prototypes of Mechatronic Systems in Virtual Environments,“ Ph.D. Thesis, University of Paderborn, HNI-Verlagsschriftenreihe, Paderborn, Vol. 193, Dez. 2006
- [16] Vance, J. and Dumont, G., „A Conceptual Framework to Support Natural Interaction for Virtual Assembly Tasks,“ in Proc. Of the ASME 2011 World Conference of Innovative Virtual Reality, WINVR2011, Jun3 27-29, Milan, Italy, 2011, pp. 1-6



## A Dance Training System that Maps Self-Images onto an Instruction Video

Minoru Fujimoto, Masahiko Tsukamoto  
*Graduate School of Engineering*  
*Kobe University*  
*Kobe, Japan*  
*minoru@stu.kobe-u.ac.jp, tuka@kobe-u.ac.jp*

Tsutomu Terada  
*Graduate School of Engineering*  
*Kobe University*  
*PRESTO, JST*  
*Kobe, Japan*  
*tsutomu@eedept.kobe-u.ac.jp*

**Abstract**—Owing to recent advancements in motion capture technologies, physical exercise systems that use human interaction technologies have been attracting a great deal of attention. There are already various approaches in place that support motion training methods by using motion capture technology and wearable sensors to analyze body motion. In this study, our basic idea was to change the appearance of a dancer in an instruction video to that of the user, who we assume would be interested in seeing what they would look like if they could perform so well. We developed a motion training system that maps the user’s image onto an instruction video. Evaluation results demonstrated that our proposed method is effective for motion training in specific situations.

**Keywords**-motion capture; motion training

### I. INTRODUCTION

Owing to recent advancements in motion capture technologies [1], human interaction technologies specialized to physical exercise [2] have attracted a great deal of attention. Motion capture technologies are popular these days, and can commonly be seen in games in which users practice dance motions. Ideally, such training systems have both a high level of motion training functionality and a high entertainment value.

There are various methods in place for analyzing body motions in motion training systems [3][4], such as image processing, motion capturing using markers, and calculating body movement by using wearable sensors. For example, Kwon et al. proposed a motion training system that uses a wireless acceleration sensor and image processing [5]. Trainers and trainees can analyze their movements by watching a hybrid representation of visual and wearable sensor data that is automatically generated as a motion training video.

In our study, the basic idea is to change the appearance of an expert dancer in an instruction video to that of the user. Users, usually, practice dance movements by following the teacher’s movements: they virtually translate each of the teacher’s body parts into their own, and consider how best to perform the movement. In contrast, with our system, we assume that players would be interested to see their dance moves as if they could already perform really well. We propose a new practice style in which a composite video is created that shows a user dancing as if he or

she were an expert. This enables users to experience what it feels like to dance skillfully. Being impressed by the performance of experts is a significant motivator for learning new techniques, so we decided to utilize this motivation for learning and support in the form of showing beginners what they would look like if they were an expert.

In the proposed motion training system, images of the user are mapped onto an instruction video. The user then practices the dance movements by following their own expertly dancing image on the video. Evaluation results demonstrated that the proposed method is effective for motion training in specific situations.

The remainder of this paper is organized as follows. In the next section we discuss related works, and in Section 3 we describe the proposed method. In Section 4, we evaluate the proposed method by training actual participants. We present our conclusions and future work in the final section.

### II. RELATED WORK

There have been many research projects that support motion training [6][7] by analyzing user motions [8] using image processing [9], motion capturing [10], and wearable sensors [11]. Chan et al. [12] proposed a virtual reality (VR)-based dance training application using motion capture technologies in which the user wears a motion capture suit and follows the movements of a virtual teacher. The system provides several types of feedback on how to improve the movements, including displaying whether the position of each body segment is correct in real-time, showing a report on the user’s performance after training, and showing a slow motion replay.

Hachimura et al. [13] proposed a motion training system using a motion capture system and mixed reality (MR) techniques. Their system uses a head-mounted display (HMD) that shows a CG character to visualize the movements of the trainer.

These methods focus on how to correct a user who has made an incorrect motion. However, users cannot imagine the situation where they are expert dancer using these methods. Our system provides the sense as if the beginner become an expert since he/she see the his/her expert performance in

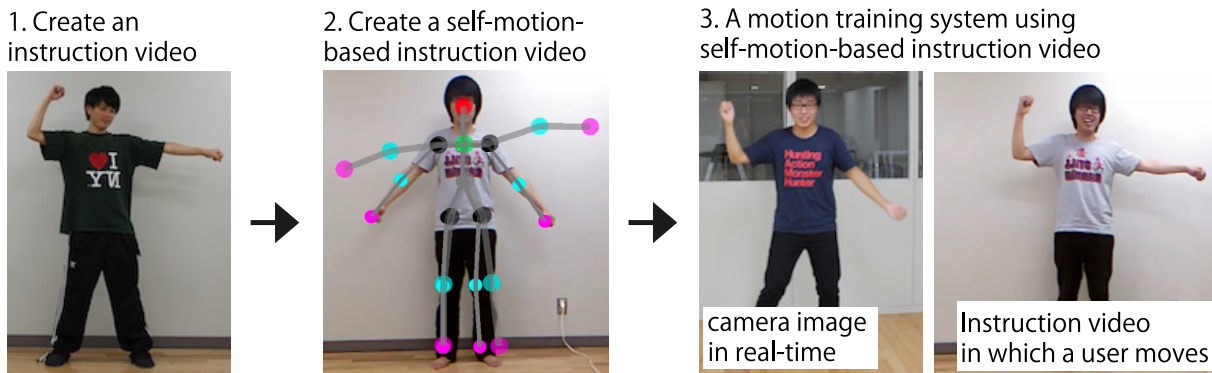


Figure 1. Procedure of the proposed system

the movie. Our method can be combined with other methods to improve its overall effectiveness.

### III. PROPOSED METHOD

As a preliminary work, we developed a dance training system that increases user motivation by using a visual effect [14]. This system vibrates the user's image by using image processing to emphasize the intensity of their movement. As many individuals experimented with our system, we noticed that it was difficult for the average person to begin to enjoy dancing. We wanted to enable such people to enjoy dancing and to master basic dance techniques, and it occurred to us that it might be effective to show them video of themselves successfully dancing.

We performed a lot of trial-and-error experiments to achieve this purpose. However, it was difficult to create a video that realistically shows the user dancing like an expert dancer because the movements are so intense. Our main goal is to examine the meaning of and effect on people who watch a video of themselves dancing well, and we decided to use one simple motion to achieve this goal: *Stop*. This is a popping motion used in street dancing. *Stop* in which a dancer stops instantaneously - like a dancing robot - in accordance with the beat of music.

#### A. System Structure

Figure 1 shows the flow of our system, which consists of three subsystems. This flow is summarized below.

- (1) The system records a video of an expert dancing to specified music while simultaneously recording the position of his/her body joints.
- (2) It plays the recorded instruction video at a very slow speed, and the user moves in accordance with the video, including the position marks of the body joints, which are used as the materials for composing the instruction video.
- (3) The user practices the dance motions by using the instruction video composed in Step (2).

We describe this process in the following sections in detail. In our prototype, we used a Kinect sensor [15] to acquire RGB and depth images. OpenFrameworks, OpenCV [16], OpenNI [17], and the NITE library were used to implement our prototype.

#### B. Recording the Instruction Video System

Our first subsystem records an instruction video that shows an expert dancing that includes RGB images, depth images, and BGM. It also records the trajectory of fifteen body joints, as shown in Figure 2(a). The frame rate for recording was 30 frames per second (fps).

#### C. Creating Self-Motion-Based Instruction Video

Our second subsystem creates a video that shows the user moving as if they are an expert dancer by using the instruction video recorded, as described in Section III.B. Figure 3 shows the flow of creating the self-motion-based instruction video. The procedure for creating the video is as follows.

- (1) The user calibrates the position of his/her body joints to that of the dancer in the instruction video.
- (2) The system plays the recorded instruction video at a very slow speed while simultaneously displaying the body joint markers.
- (3) The user moves in accordance with the markers in the instruction video and the system records the user's images and the position of the body joints on each frame.
- (4) The system calculates the distance between the position of the recorded user's body joints and that of the corresponding body joints on the instruction video on each frame and then determines the user's image for each frame of the instruction video.

This calibration process is described in detail in the following subsection.

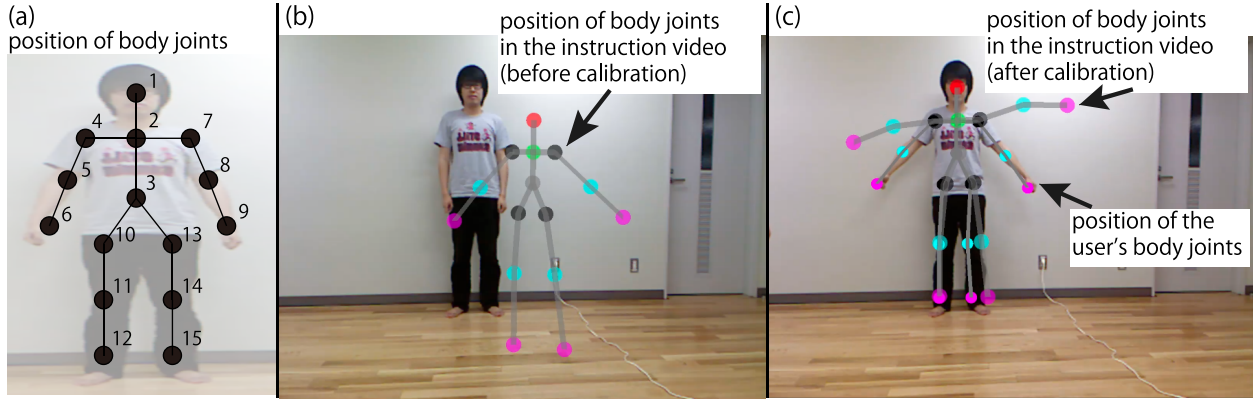


Figure 2. A calibration technique to position body joints

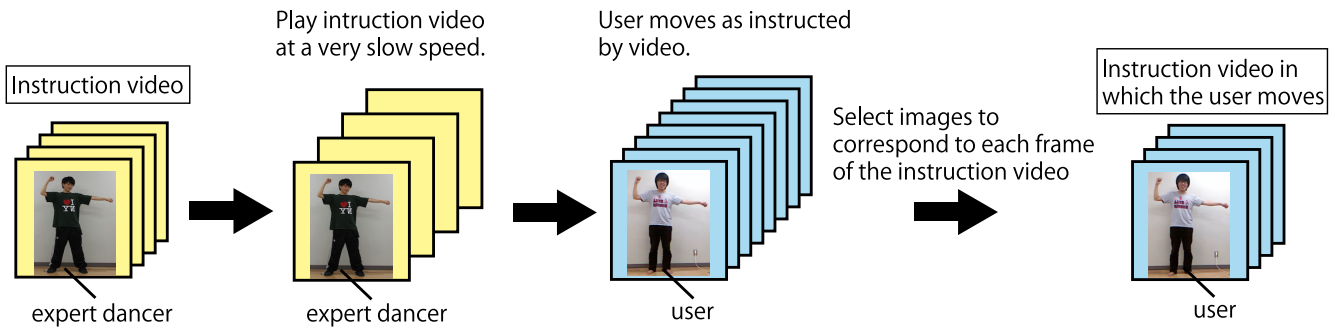


Figure 3. Procedure of creating a self-motion-based instruction video

1) *Calibration of Body Joints:* We use fifteen body joints, as shown in Figure 2(a). If the system displays the position of body joints from the instruction video without any modification, the user cannot move in accordance with it due to the difference between the size of the user's and dancer's bodies (Figure 2(b)). Therefore, our system calculates the distance between the user's body joints and the body joints on the instruction video and displays the body joints on the instruction video as adjusted to those of the user.

At a given number  $n$ , the system defines the body joints on the instruction video,  $S_n(x, y)$ , and the user's body joints,  $U_n(x, y)$ . The Euclidean distance  $d(S_{n1}, S_{n2})$  and  $d(U_{n1}, U_{n2})$  are calculated as:

$$d(S_{n1}, S_{n2}) = \sqrt{(S_{n1} - S_{n2})^2}$$

$$d(U_{n1}, U_{n2}) = \sqrt{(U_{n1} - U_{n2})^2}$$

The user can perform the same movements as the expert dancer by fitting the position of their joints to these on the instruction video,  $S'_n(x, y)$ , calculated as follows.

First, the system adjusts the center of the dancer's body joints,  $S_3(x, y)$ , to the center of the user's body joints,  $U_3(x, y)$ .

$$S'_3(x, y) = S_3(x, y) + \{U_3(x, y) - S_3(x, y)\} = U_3(x, y)$$

Next, the system adjusts  $S_2(x, y)$  to  $U_2(x, y)$  and calculates a ratio of length by using  $d(S_2, S_3)$  and  $d(U_2, U_3)$ . In addition, because  $S'_3(x, y)$  is adjusted to  $U_3(x, y)$ ,  $S'_2(x, y)$  needs to adapt to the difference  $\{U_3(x, y) - S_3(x, y)\}$ .

$$S'_2(x, y) = S_2(x, y) + \{U_3(x, y) - S_3(x, y)\} \\ + (1 - \frac{d(U_2, U_3)}{d(S_2, S_3)}) \{S_3(x, y) - S_2(x, y)\}$$

$S'_4(x, y)$  is calculated by using the result of  $S'_3(x, y)$  and  $S'_2(x, y)$ :

$$S'_4(x, y) = S_4(x, y) + \{U_3(x, y) - S_3(x, y)\} \\ + (1 - \frac{d(U_2, U_3)}{d(S_2, S_3)}) \{S_3(x, y) - S_2(x, y)\} \\ + (1 - \frac{d(U_2, U_4)}{d(S_2, S_4)}) \{S_2(x, y) - S_4(x, y)\}$$

In the same way, the system calculates all the positions of the body joints on the instruction video that have been adjusted to the user's body joints. As shown in Figure 2(c), the adjusted positions of the body joints on the instruction video overlap the user's body.

2) *A Recording Technique Using Slow Instruction Video:* The system creates a composite video showing the user dancing like an expert by recording the user's movements

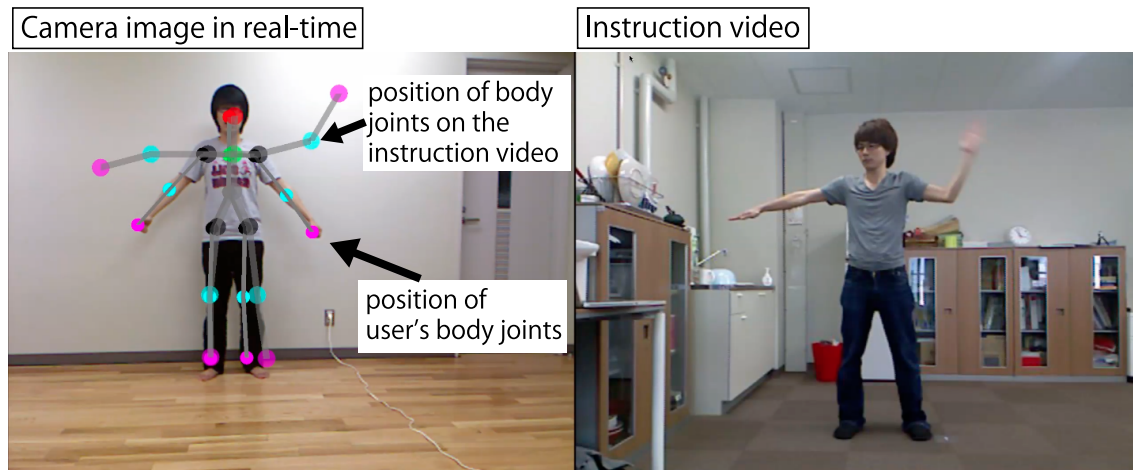


Figure 4. Adjustment technique for body joints

in accordance with the instruction video. Obviously, it is difficult for beginners to dance like experts, so the system plays the instruction video at a slow speed, giving the user a longer time to get the movements right.

Specifically, the system plays the instruction video at a speed of one fifth and displays the adjusted position of body joints  $S'_n(x, y)$ . As shown in Figure 4, the system simultaneously displays the position of the user's body joints and the user follows the adjusted positions of the body joints on the instruction video. As the system plays the instruction video at the slower speed, it acquires the user's RGB image, depth image, and position of body joints at 20 fps. Therefore, when the system creates the composite video for 10 seconds, the user follows the instruction video for 50 seconds and the system records 1000 images at 20 fps. To summarize: the system creates a composite video showing the user dancing expertly from 1000 images.

3) *Creating a Composite Self-Motion Image*: As described in the previous section, the system records many images of the user, which it then uses to create the composite video. The sum of the Euclidean distance between the position of the body joints on the instruction video  $S'_n(x, y)$  and the position of the user's body joints  $U_n(x, y)$  is calculated as:

$$d(S', U) = \sqrt{\sum_{k=1}^{15} (S'_k - U_k)^2}$$

The system determines each frame of the composite video as the frame that has a minimum  $d(S', U)$ . By playing the images of the user that correspond to each frame of the instruction video at 30 fps, we can create a video that shows the user dancing like an expert.

#### D. Practicing with the Composite Self-Motion Image

The user can practice dance movements by viewing a real-time camera image on the left side and the instruction video on the right side, as shown in Figure 4. The user can switch between the instruction video and the composite video, which enables him/her to select the most effective video for whatever he/she is practicing. We also implemented a function that records a practice video and displays it on the left side, so the user can check his/her movements against the instruction video simultaneously.

## IV. EVALUATION

Our proposed system enables users to view an imitation of themselves dancing like an expert. In this section, we evaluate the effectiveness of the proposed system. The subjects used for each evaluation were 15 college students (including 1 expert dancer).

First, we created the dancer's instruction video. This video was 4 seconds long and featured a dancer performing *Stop* - in which the dancer pops like a robot in accordance with the beat - at 120 beats per minute. The movements included two types of motion: one for 250 ms at 750-ms intervals and the other for 250 ms at 250-ms intervals.

To create the composite video, subjects followed the instruction video at a speed of one fifth, as shown in Figure 4. Next, the subjects practiced the movements by following the instruction video for 3 minutes, as shown in Figure 5(a). Next, the subjects recorded their movements and checked them against the instruction video for 5 minutes, as shown in Figure 5(b). In each phase, the subjects were able to switch between their own video and that of the expert dancer. The subjects then answered the following three questions:

- (1) If you practiced the movements by following the instruction video, which did you prefer, the instruc-

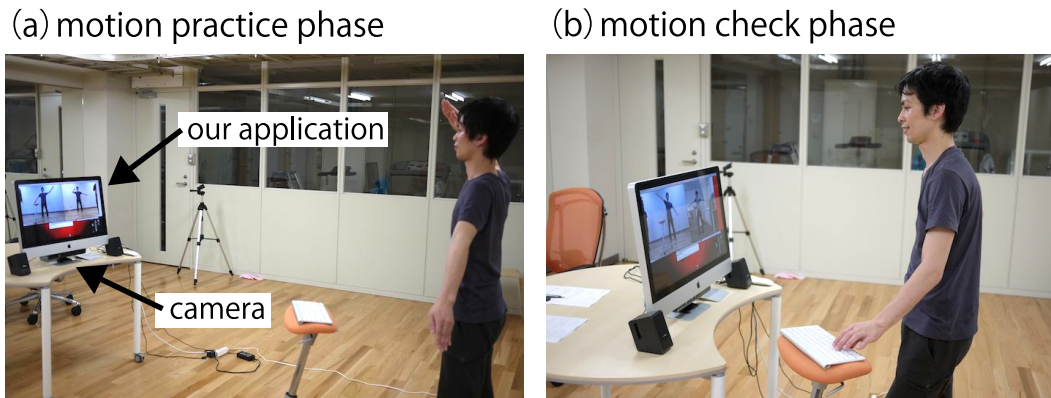


Figure 5. Snapshots of evaluation

Table I  
EVALUATION RESULTS

Evaluation score	Number of subjects					The meaning of each evaluation
	1	2	3	4	5	
Evaluation (1)	3	3	3	5	1	1 (dancer's video) - 5 (user's composite video)
Evaluation (2)	4	1	3	3	4	1 (dancer's video) - 5 (user's composite video)
Evaluation (3)	0	0	0	5	10	1 (not interesting) - 5 (interesting)

tion video of the dancer or your own composite instruction video?

- (2) If you recorded your movements and checked them against the instruction video, which did you prefer, the instruction video or the composite video?
- (3) Do you think this system is interesting?

Table I lists the results of the evaluation. In questions (1) and (2), the opinions of the subjects are remarkably divided. We describe the evaluation and our conclusions in the following section.

#### A. Evaluation 1: Motion Practice Phase

The responses to question (1) are as follows.

- The composite video of myself was good. It reflected my movements more than the dancer's instruction video.
- I had a sense of rhythm by using the composite video. But I prefer the dancer's instruction video for practicing specific details of the movements.
- I was able to imagine myself successfully dancing by watching the composite video.
- I watched the dancer's instruction video more than the composite video of myself, because I felt that the instruction video was more accurate.
- First, I watched my composite video. Next, I watched the dancer's instruction video. This flow was effective for learning the movements.
- I disliked my composite video because it embarrassed me.

- I was skeptical of my composite video because I can't dance.

For subjects who were able to imagine themselves as dancers, it was effective to watch the composite video. Subjects who preferred the instruction video had the pre-conception that the instruction video was more correct than the composite video. Furthermore, they were not able to completely trust their own composite videos. Some subjects had different video preferences for each phase of learning the movements. We need to examine the role of timing in the use of the composite video.

#### B. Evaluation 2: Motion Check Phase

The responses to question (2) are as follows.

- It was helpful for me to study parts of my body by using the composite video.
- Using the composite video was helpful in terms of timing the movements.
- I was mortified when I was not able to dance well by using my composite video. This motivated me to try harder.
- By using the composite video, I could see my mistakes clearly.
- First, I watched my composite video. Next, I watched the dancer's instruction video. This flow was helpful for me to check the precision of my dancing.
- I preferred using the dancer's instruction video for checking my mistakes.
- Watching the dancer's instruction video made me feel that I have not yet mastered the dance.

In the phase of checking movements with the instruction video, subjects were able to compare the features of their bodies more than during the practicing phase. Subjects who preferred the dancer's instruction video used it to check the quality of their movements, while those who preferred the composite video used it to check the timing of their movements. In other words, preference for the dancer's instruction video versus the composite video differed depending on which movements were being learned. We intend to further evaluate our system to clarify the timing for using the composite instruction video in the future.

### C. Evaluation 3: Impression of Subjects

The responses to question (3) are as follows.

- It was interesting to create a composite video by moving slowly.
- When I was able to dance well, my movements resembled my composite video. It was very interesting.
- By watching my composite video, I felt that I was able to dance well.
- My composite video was a little awkward.
- I felt a slight sense of incongruity in my composite video, but I still felt that it is effective for learning the movements.

By using our system, the subjects were able to enjoy learning dance movements. It seems that clear goals are effective when a beginner first learns movements. However, we need to improve our system to create a composite video of higher quality.

### V. CONCLUSION

We proposed a system that enables beginners to practice dance movements by studying self-motion images that have been mapped onto an instruction video. Our system creates a composite video in which beginners appear to dance like experts. Experimental results showed that our system is effective for motion training in specific situations.

However, the motions that can be used are limited, and the composite video is not of high quality. We intend to improve the system in these respects in the future.

### ACKNOWLEDGMENT

This research was supported in part by a Grant in aid for Precursory Research for Embryonic Science and Technology (PRESTO) from the Japan Science and Technology Agency and by a Grant-in-Aid for Scientific Research(A)(20240009) from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

### REFERENCES

- [1] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake: Real-time human pose recognition in parts from single depth images, *Proceedings of Computer Vision and Pattern Recognition (CVPR'11)*, pp. 1297-1304, 2011.
- [2] P. Haemaelaenen, T. Ilmonen, J. Hoeynsniemi, M. Lindholm, and A. Nykaenen: Martial arts in artificial reality, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 781-790, 2005.
- [3] U. Yang, and G. J. Kim: Just Follow Me: An Immersive VR-based motion tracking system, *Proceedings of the International Conference on Virtual Systems and Multimedia*, pp. 435-444, 1999.
- [4] U. Yang, and G. J. Kim: Implementation and evaluation of "Just Follow Me": an Immersive, VR-based, motion tracking system, *Presence: Teleoperators and Virtual Environments*, MIT Press, Vol. 11, No. 3, pp. 304-323, 2002.
- [5] D. Y. Kwon and M. Gross: Combining body sensors and visual sensors for motion training, *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pp. 94-101, 2005.
- [6] J. N. Bailenson et al: The effect of interactivity on learning physical actions in virtual reality, *Media Psychology*, Vol. 11, pp. 354-376, 2008.
- [7] S. Baek, S. Lee, and G. Kim: Motion retargeting and evaluation for VR-based training of free motions, *The Visual Computer*, Vol. 19, pp. 222-242, 2003.
- [8] T. Shiratori, A. Nakazawa, and K. Ikeuchi: Rhythmic Motion Analysis using Motion Capture and Musical Information, *Proceedings of 2003 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems(MFI'03)*, pp. 89-94, 2011.
- [9] L. Naveda and M. Leman: Representation of Samba dance gestures, using a multi-modal analysis approach, *Proceedings of Fifth International Conference on Enactive Interfaces*, pp. 68-74, 2008.
- [10] C. Chua, N. H. Daly, V. Schaaf, and H. P. Camill: Training for physical tasks in virtual environments: Tai chi, *Proceedings of IEEE Virtual Reality 2003 Conference*, pp. 87-94, 2003.
- [11] K. Matsumura, T. Yamamoto, and T. Fujinami: Analysis of Samba Dance Using Wireless Accelerometer, *Nihon Kikai Gakkai Supotsu Kogaku Shinpojiumu, Shinpojiumu Hyuman, Dainamikusu Koen Ronbunshu*, Vol. 2006, pp. 216-221, 2006.
- [12] J. Chan, H. Leung, J. Tang, and T. Komura: A virtual reality dance training system using motion capture technology, *IEEE Transactions on Learning Technologies*, Vol. 4, pp. 187-195, 2010.
- [13] K. Hachimura, H. Kato, and H. Tamura: A Prototype Dance Training Support System with Motion Capture and Mixed Reality Technologies, *Proceedings IEEE Int'l Workshop Robot and Human Interactive Comm*, pp. 217-222, 2004.
- [14] A/D Dance, <http://ivrc.net/2010/en.html>, 01.12.2011
- [15] Microsoft XBOX Kinect, <http://www.xbox.com/en-US/Kinect>, 01.12.2011
- [16] OpenCV <http://opencv.jp/>, 01.12.2011
- [17] OpenNI, <http://www.openni.org/>, 01.12.2011

# Interaction in Augmented Reality by Means of Z-buffer Based Collision Detection

Yasuyuki Souma, Hidemi Yamachi, Yasuhiro Tsujimura, Yasushi Kambayashi  
 Department of Computer and Information Engineering,  
 Nippon Institute of Technology 4-1 gakuendai,  
 Miyashiro-cho, Minamisaitama-gun, Saitama 345-8501 Japan  
 c1065309@cstu.nit.ac.jp, yamachi@nit.ac.jp, tujimr@nit.ac.jp, yasushi@nit.ac.jp

**Abstract**— We propose a new interaction method in AR (augmented reality) using a depth sensor and a collision detection method with the Z-buffer. The method generates 3D models of the real world in real-time and performs interaction between the real and virtual world. In order to provide the real-time collision detection between the models of real world and those of virtual world, we have developed a detection method called *Cyber Radar*. This technique uses two sets of depth values of the Z-buffer that is generated through the orthographic projection. One is obtained through rendering the virtual space from the sensor object toward a target point. This set does not have the depth values of the sensor object. Another one is obtained through rendering only the sensor object in the reverse direction. From these two depth value sets, we obtain the distance between the sensor object and others for each pixel. This technique requires only one or two rendering processes and it is independent from the complexity of object's shape, deformation or motion. In this paper, we evaluate the interaction using this method and report the problems to be solved.

**Keywords**—*Cyber Radar; depth sensor; Z-buffer; collision detection; object detection*

## I. INTRODUCTION

For detecting and tracking objects in augmented reality, people use techniques that are based on the recognition of markers or patterns [1]. The virtual objects make action in the relation of those markers. When some other objects hide the markers, the virtual objects can no longer exist in the virtual world [2]. It is impossible for those objects to move around in the virtual and interact with any object in the real world without markers. In addition to that, since the virtual objects are projected on the real world scene, they are never overridden by real objects. These problems of AR techniques give us unnatural feelings when we see the augmented reality scenes.

Hence we need to generate 3D models of the real world to interact virtual objects with real objects in real-time. There are some techniques for generating 3D models of real world. Recently Microsoft released a motion sensing input device named Kinect [3] for their video game machine. Kinect features an RGB camera and a depth sensor. It provides us the depth values of the scene in real-time. Using these depth values, we can construct the 3D models of the real world. Since the models of the real world are just for computing interaction, not for displaying the objects in the real world, it

is not necessary to generate exact 3D models with expensive calculation cost.

In order to generate the interaction between the virtual objects and the 3D models of real world, we need to adopt a collision detection method that does not depend on the number of objects or shapes and motions of them.

The problem of collision detection in cyberspace has been intensively studied in various fields. Most of them have the same feature. They calculate geometrical intersections between each object pair. For reducing the calculation cost, they divide the space or objects into convex polygons and eliminate obviously not colliding parts. The first type of techniques divides whole space into cells and keeps positions of objects using hierarchical trees, such as octree [4], BSPtree [5] and  $k$ -d trees [6]. The second type of techniques uses bounding volume hierarchies (BVH) which consist from convex hulls such as spheres or boxes. Convex hulls hierarchically cover an object from the whole to each part. AABB [7], Bounding Sphere [8], OBB [9],  $k$ -DOPs [10] are well known and used for collision detection. Because those techniques detect collisions only one pair of objects at once, they have  $O(n^2)$  computational complexity for  $n$  objects. They work very well for rigid models that never change their shapes. Once the models change their shapes, however, they have to reconstruct the tree or BVH. The reconstruction procedure takes considerable time and space computational complexities. All of those methods perform one collision test for every pair of objects. If there are several potentially colliding object pairs, they have to perform the collision tests for all of them.

**Main contribution:** We propose an application of the interactive AR system using a depth sensor and the collision detection method, named *Cyber Radar* [11]. In [12], we showed *Cyber Radar* with the latest hardware showed  $O(n)$  performance for  $n$  polygons and reading back Z-buffer values from GPU memory is not the bottle neck for our method. *Cyber Radar* detects collisions between each predefined object and other objects. We call the object *sensor object*. *Cyber Radar* uses two sets of Z-buffer values obtained through rendering with the orthographic projection. The viewing volume for the rendering is defined to include the space in which collisions could occur. We also call the viewing volume *target volume*. At first, we set the view point at the sensor object and render the scene toward the end of the viewing volume without the sensor object. If there is no depth value in the Z-buffer less than the maximum depth value, no other object in the target volume and no

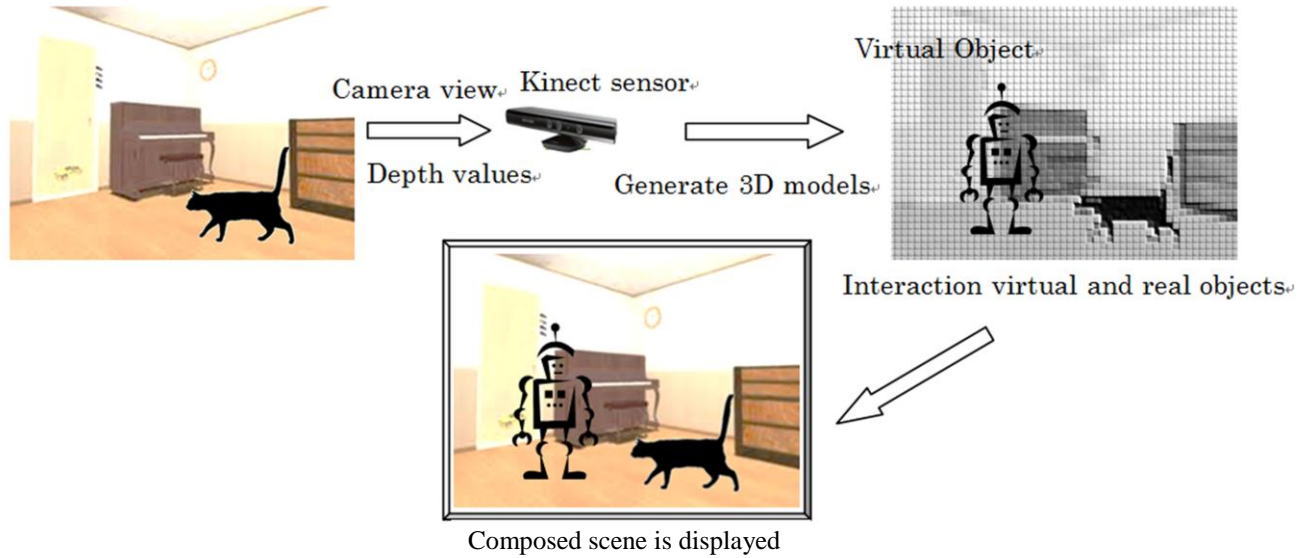


Figure 1. Structure and procedure of the AR system

collision will occur. Otherwise, we set the view point at the opposite position and render only the sensor object, without any other objects. From those two sets of depth values, we can find the distance between the sensor object and other objects in the target volume.

*Cyber Radar* takes only two renderings for one collision test. Different from other methods, it detects multiple collisions at once without selecting any object pair. Then *Cyber Radar* performs collision detection almost  $O(n)$  time, where  $n$  is the number of polygons. *Cyber Radar* is also independent from the complexity of an object's shape, deformation or motion. In addition to these advantages, the implementation is simple and easy.

Because of these features of *Cyber Radar*, it is natural for us to apply it for detecting collisions in the AR system with the depth sensor. We have implemented our AR system with *Cyber Radar* and evaluate its effectiveness and investigate problems to be solved.

**Organization:** The rest of this paper is organized as follows. In the next section, we describe the structure of our AR system. In section III, the algorithm of *Cyber Radar* is described. We show the efficiency and problems of our method through experiments in section IV. We conclude our discussions in section V.

## II. SYSTEM STRUCTURE

Our AR system uses Kinect sensor to obtain the RGB camera view and the depth values of the view. Then the system generates the 3D models of the view. After calculating the interaction between the real and the virtual objects, the camera view and the virtual image are superimposed on the PC display. The overall structure and its procedures are depicted in Figure 1.

Aiming for realizing our AR system with real-time interaction, we need to generate 3D models of the real world

without intensive computation. Then our AR system generates models of the real world by arranging polygons for each pixel at each corresponding coordinate in the cyberspace.

## III. MECHANISM OF CYBER RADAR

In order to detect other objects in the target volume, we set the view point at the sensor object (Figure 2(a)). By defining the orthogonal view volume that has the same depth as the length of the sensor object and the distance to detect collision, as the target volume, the scene is rendered without the sensor object (Figure 2(b)). We call the obtained the Z-buffer value set *Range Distance Scope* denoted as  $\mathbf{R}$ . If there are any pixels in the *Range Distance Scope* that have the depth values less than the target volume distance  $d$ , there are some objects in the target volume.

If some objects are detected in the target volume, the collision detection is performed. The view point is set at the other side of the target volume with the direction toward the sensor object. Then only the sensor object is rendered to obtain the depth value set. We named the Z-buffer value set *Mask Distance Scope* denoted as  $\mathbf{M}$ . With these two depth value sets, we can obtain the distances between the sensor object and other objects. We named this distance value set *LookAt Distance Scope* denoted as  $\mathbf{L}$ .

Before we describe the detail algorithm of *Cyber Radar*, we define the following notations.

- $p_w, p_h$  : Pixel width and height of depth values  
 $i = 1, 2, \dots, p_w, j = 1, 2, \dots, p_h$
- $\mathbf{R}$  : *Range Distance Scope*  $\mathbf{R} \in \{R_{ij}\}$
- $\mathbf{M}$  : *Mask Distance Scope*  $\mathbf{M} \in \{M_{ij}\}$
- $\mathbf{L}$  : *LookAt Distance Scope*  $\mathbf{L} \in \{L_{ij}\}$
- $d$  : Target volume distance



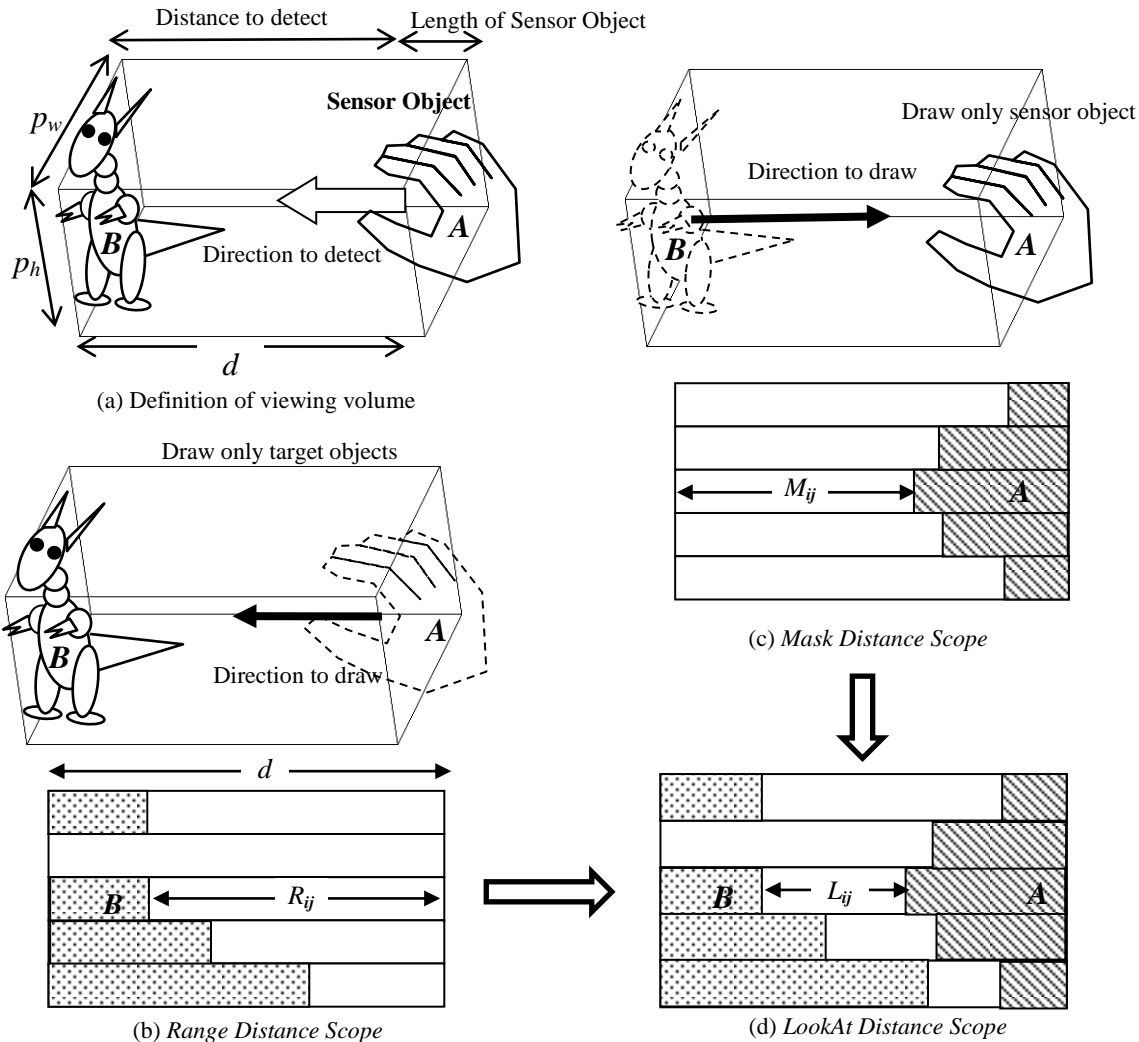


Figure 2. Collision detection by Cyber Radar

The distance between view point and the detected objects is  $d - R_{ij}$  and the distance from the end of the target volume to the sensor object is  $d - M_{ij}$ . Using these values, the pixel value of *LookAt Distance Scope* becomes as follows.

$$L_{ij} = d - (d - R_{ij}) - (d - M_{ij}) = R_{ij} + M_{ij} - d \quad (3)$$

If  $\min(L)$ , the minimum value of  $L$ , is greater than zero or distance to move, collision will not occur. Otherwise collision will occur at the pixel(s) that have the minimum value. If objects are moving, the time of collision is calculated from  $\min(L)$  and the velocity of objects.

The algorithm of *Cyber Radar* is described as follows.

1. Set up the rendering conditions (Figure. 2(a))
  - Set view point at the sensor object
  - Determine rendering direction
  - Set width and height of the target volume
  - $d :=$  Target volume distance
  - $p_w, p_h :=$  Projection width and height
2. Render scene without the sensor object (Figure. 2(b))

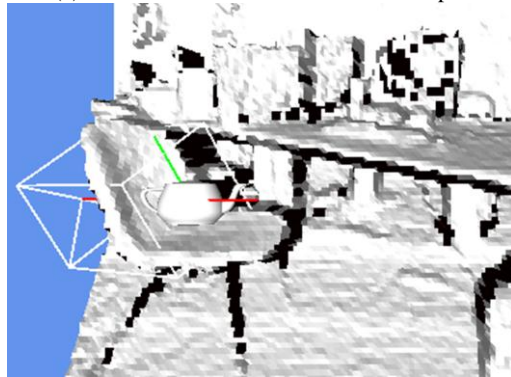
3. Obtain  $\mathbf{R}$  (Range Distance Scope)
4. **if**  $\min(\mathbf{R}) = d$ 
  - return false** (no object)
5. Set the rendering conditions (Figure 2(c))
  - Set view point at the end of the target volume
  - Determine rendering direction toward sensor object
6. Render scene only sensor object
7. Obtain  $\mathbf{M}$  (*Mask Distance Scope*)
  - Swap right and left pixel values
8. Calculate  $\mathbf{L}$  (*LookAt Distance Scope*)
  - $\mathbf{L} := \mathbf{R} + \mathbf{M} - d$
9. **if**  $\min(\mathbf{L}) > 0$  or distance to move
  - return false** (no collision)
  - else**
  - return true** (collision is detected)

#### IV. IMPLEMENTATION RESULTS AND ESTIMATION

We have implemented the AR system to estimate its performance. Figure 3(a) shows the scene that the real world with the virtual object. The teapot is the virtual object. Figure



(a) Scene of real world and virtual tea pot



(b) 3D models of the real world and virtual teapot

Figure 3. The scene of the AR system

3(b) shows the 3D models of the same scene that is generated to detect collision between the real objects and the virtual teapot. The white rectangular frame at the teapot is the target volume of *Cyber Radar*. The resolution of the depth sensor of Kinect is 640x480 pixels. In order to generate models of the real world, we put polygons at each coordinate of pixels by considering their real sizes and depth values. The number of polygons is 37,842 for every scene of the models of the real world while 6, 6,256 and 181,412 are for three sensor objects. The computing environment for the experiment is Mac Pro with Quad-Core Intel Xeon 3500 2.66GHz and NVIDIA GeForce GT 120.

The frame rates with collision detection by *Cyber Radar* are in Table 1. As the number of polygons of the sensor objects increases, the frame rate decreases noticeably. According to our ongoing experiments, it seems that the modeling procedure causes the low frame rate. We need to analyze the bottle neck of the system exactly and devise the solution for increasing the frame rate.

## V. CONCLUSION

We proposed a new interaction method in AR using a depth sensor and a collision detection method using Z-buffer named *Cyber Radar*. *Cyber Radar* detects collision independent from the complexity of object's shape, deformation or motion. Through the experiment, our AR system demonstrates that it achieves the required performance for real-time interaction. Through the experiment, we found some problems to be solved.

Table 1. Frame rate with collision detection

Number of polygons of sensor object	fps
6	13.9
6,256	12.5
181,412	9.4

There may be hidden surfaces in each scene of the real world. There are no polygons at those parts. If the virtual object moves into those parts, it falls into behind the real world. To avoid this problem, we may need to add polygons at the edges of those parts toward the polygons of their neighbor pixels. This means that closer objects in the real world are embossed.

To solve this problem, we need to develop the method for modeling of the real world. While static objects in the real world can be modeled by scanning with the depth sensor in advance, we need to generate models of dynamic objects in real-time.

We expect to reduce the number of polygons of the virtual objects for rendering according to the resolution of the depth sensor. We also expect that this method reduces the computation time for collision detection.

## ACKNOWLEDGMENT

This work is partially supported by Japan Society for Promotion of Science (JSPS), with the basic research program (C) (No. 23510178), Grant-in-Aid for Scientific Research.

## REFERENCES

- [1] G.Klein and D.Murray, "Parallel Tracking and Mapping on a camera phone," ISMAR 2009. Proceedings of 8th IEEE International Symposium on Mixed and Augmented Reality, pp. 83-86 2009
- [2] H.Uchiyama and H.Saito, "Random dot markers," IEEE Virtual Reality Conference 2011, pp. 35-38 2011
- [3] <http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/> <retrieved: September, 2011>
- [4] H. Samet, "Spatial Data Structures: Quadtree, Octrees and Other Hierarchical Methods," Addison Wesley 1989
- [5] B. Naylor, J.A. Amatodes and W. Thibault, "Merging BSP trees yields polyhedral set operations," Computer Graphics (SIGGRAPH '90 Proc.), vol. 24, pp. 115-124 1990
- [6] M. Held, J.T. Klosowski and J.S.B. Mitchell, "Evaluation of collision detection methods for virtual reality flythroughs," Proc. 7th Canada Conf. Comput. Geom., pp. 205-210 1995
- [7] M. Ponangi, D. Manocha and M. Lin, "Incremental algorithms for collision detection between solid models," IEEE Transactions on Visualization and Computer Graphics, vol.3(1), pp. 51-67, 1997.
- [8] P.M. Hubbard, "Approximating Polyhedra with Spheres for Time-Critical Collision Detection", ACM Transactions on Graphics, vol. 15(3), pp. 179-210, 1996.
- [9] S. Gottschalk, M. Lin and D. Manocha, "OBBTree: A Hierarchical Structure for Rapid Interference Detection", Proc. of ACM Siggraph '96, pp. 171-180 1996
- [10] G. van den Bergen, "Efficient Collision Detection of Complex Deformable Models using AABB Trees." Journal of Graphics Tools, vol.2(4), pp. 1-14 1997
- [11] H. Yamachi and Y. Shindo, "A Technique for Object and Collision Detection by Z-buffer," Transactions of Information Processing Society of Japan vol. 43(6), pp. 1899-1909, 2002 (in japanese)
- [12] H.Yamachi, Y.Souma, Y.Kambayashi, Y.Tsujimura and T.Iida, "Evaluation of a Technique for Collision and Object Detection with the Z-buffer in Cyber Space," Cyberworlds 2011, pp. 26-92 2011

## The Effect of Metacognition in Cooperation on Team Behaviors

Kohei Nonose, Taro Kanno, and Kazuo Furuta  
 The University of Tokyo  
 Tokyo, Japan.  
 nonose@cse.sys.t.u-tokyo.ac.jp

**Abstract**—Teams and teamwork are indispensable, especially when tackling difficult and complex tasks that cannot be easily addressed by a single individual. Because breakdowns in team cooperation can cause accidents, much research attention has been devoted to studies on team cooperation, and many measurements and training of teamwork have been proposed. Traditional studies have often focused on observational teamwork behaviors to measure and enhance teamwork. In order to better measure and enhance teamwork, it is believed that it is necessary to focus on the cognitive mechanisms that underlie teamwork. This study focuses on metacognition in cooperation that underlies team cooperation, and aims to investigate the importance of metacognition in cooperation. The comparisons of metacognition in cooperation and team performance indexes suggest that an improvement of metacognition in cooperation will enhance team performance and that certain types of metacognition in cooperation are important for positive teamwork.

**Keywords** - team cooperation, cognition, measurements, training.

### I. INTRODUCTION

Team performance has been increasingly recognized as an indispensable foundation of difficult and complex tasks that cannot be easily addressed by a single individual, such as air traffic control and surgical care. The ability of an individual to contribute as the member of a team in a complex team task should be enhanced through training strategies that are aimed at providing competencies that facilitate teamwork. One of the prerequisites of team training is valid and reliable teamwork measurement, which underlies effective team performance. This is because it is necessary to identify the problems and characteristics of a team, provide constructive feedback, and evaluate the success of training for the training to be successful.

One of the typical measurements of teamwork is based on the behavioral marker system of teamwork. Mishra et al., for example, have developed a measurement of non-technical skills (NOTECHS), including teamwork [1]. They have divided NOTECHS into four behavioral dimensions (leadership and management, teamwork and cooperation, problem-solving and decision-making, and situation awareness) and defined the positive/negative behavioral modifiers of these dimensions. Based on these modifiers, trained experts evaluate NOTECHS using a four-point scale. Crew Resource Management (CRM) has been proposed and adopted as a procedure of teamwork training in different industries and organizations. CRM focuses on improving

teamwork behaviors, including interpersonal communication, leadership, and decision-making [2].

The traditional research and methods described above have often focused on observational teamwork behaviors to measure and enhance teamwork. Explicit teamwork behaviors can be beneficial for the assessment of teamwork; however, it is additionally necessary to focus on both implicit teamwork and the cognitive aspects of teamwork in order to better measure teamwork. An improvement of the cognitive mechanisms underlying teamwork must effectively be able to enhance team performance. A recent study has implied that an important mechanism behind team cooperation is metacognition in cooperation [3]. This study aims to show the effect of metacognition on cooperation in effective team behaviors and to verify useful metacognition in cooperation for positive teamwork. The next Section introduces a team cognition model about metacognition in cooperation. In Section III, a team experiment and an analysis of the elicited reflection are described. In Section IV, the analysis and team performance indexes are compared. In Section V, we conclude this study.

### II. A TEAM COGNITION MODEL

Cognition in teams has been receiving much research attention for more than a decade, and a variety of cognition models for teams have been proposed. Many of these models have aimed to present either the status of cognition in teams or the sum/overlap of individual cognition [4], instead of describing the cognitive factors underlying cooperation. We examined these factors through participants' reflections on cooperation, and proposed a team cognition model that describes and explains the cognitive processes of cooperation [3]. As a method to examine the underlying cognitive factors of cooperation, we analyzed participants' reflections on cooperation and elicited several important factors of cooperation. Subsequently, based on these factors, the findings of past studies by team researchers, and human cognitive abilities [5][6][7], we developed a team cognition model to capture a portion of the cognitive factors of team cooperation. The schematic of the proposed model is shown in Fig 1. The model consists of two levels—object-level and meta-level—with reference to a popular structure of metacognition that is defined as “cognition about cognition” [5]. The object-level is the ongoing progress or current status of a particular cognitive activity and is described by the combination of

two categories—“Subjects” and “Contents”. Object-level is monitored and controlled by meta-level, and the abilities for this process are defined as “Metacognitive skills”. The details of each category are shown in Table I and are described in the following subsections.

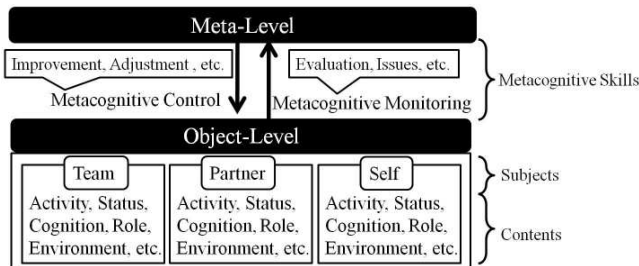


Figure 1. A Team Cognition Model

A. Subjects

Reviews and studies about group intention imply that humans use two modes of perspective in collective activities: reductive and non-reductive [8]. In the non-reductive perspective mode, one recognizes his or her group as having a mind of its own and this is distinct from individual members’ intentions. The notion of “group mind” represents this mode [6]. In the reductive perspective mode, on the other hand, the individual infers or simulates the minds of others, consciously or subconsciously, and relates this to his or her own intention and the inference results. The notion of “mutual belief” represents this mode [7]. This category of the model presents the two modes of perspective in collective activities, and consists of three elements: “Self,” “Partner,” and “Team”. “Self” and “Partner” represent the reductive perspective, and “Team” represents the non-reductive perspective.

B. Contents

This category represents content that relates to operations and consists of a variety of mental processes, actions, equipment conditions, situation awareness, mental status, knowledge, roles, tactics, etc. A set of one of the subcategories of this category and that of “Subject” comprises a basic element of the object-level. Non-cognitive elements such as equipment and roles were reported in introspections on team cooperation; thus, these elements are also included in this category.

C. Metacognitive Skills

Metacognitive skills are defined as monitoring and controlling the object-level (the set of “Subjects” and “Contents”) through the meta-level. This category helps represent cognitive factors behind teamwork behaviors and teamwork behaviors themselves, in combination with the object-level. Examples are shown in the subsequent sections.

D. Characteristics

An advantage of the proposed model is that it describes the cognitive aspects of team cooperation and explains the reason behind popular teamwork as a set of “Metacognitive Skills,” “Subjects,” and “Contents”. For example, mutual performance monitoring that is considered important teamwork behavior [9] can be described as applying “Evaluation (Good/Bad)/Issue” to “Activity/Status” of “Self/Partner/Team”; the shared mental model of role sharing can be described as applying “Compare (Match)” to the “Role” of “Team”; and behavior adjustments can be described as applying “Adjustment” to “Activity” of “Self”.

Our previous study has implied that a wide scope of metacognition in cooperation will provoke metacognitive skills in both team members’ activities and interactions; this could encourage team members to strive to improve their team performance. However, the importance of metacognition in cooperation, especially metacognitive skills for cooperation, has not completely been clarified.

TABLE I. THE CATEGORIES OF THE MODEL

Category	Subcategory	Explanation
Subjects	Self	The subject of content is oneself.
	Partner	The subject of content is a partner.
	Team	The subject of content is a team.
Contents	Activity	Actual actions, activities, decision-making activities, communication, etc.
	Cognition	Perception, comprehension, prediction, and thought.
	Tactics	Principle, operating procedures, tactics, etc.
	Role	Role sharing in a team.
	Status	Performance and workload.
	Environment	Environmental conditions regarding task accomplishment (e.g., equipment, positions of members).
Metacognitive Skills	Adjustment	Making modifications.
	Compare (Match/Mismatch)	Recognizing agreement/disagreement with the partner.
	Belief in Partner	Thinking about what his or her partner is thinking about.
	Improvement	Ideas for improving “Content.”
	Issues	Pointing out problems.
	Clear/Not Clear	Content is clear/not clear.
	Evaluation (Good/Bad)	Giving some content a good/bad evaluation.
	Characteristics	Understanding characteristics of the team environment and task rules.

If metacognition in cooperation is important for positive teamwork, the reflection in cooperation by team members who demonstrate good team performance must include a variety of subjects, contents, and metacognitive skills. This study aims to test this hypothesis and investigate the importance of metacognition in cooperation.

III. EXPERIMENT AND DATA ANALYSIS

In order to investigate the importance of metacognition in cooperation, we use the experiment data collected in [3] and analyze the relationship between the coded results of the reflection on cooperation and team performance in detail. First, the team experiment is introduced. Then, the coded results of the reflection on cooperation are shown. Finally, the reflection with the team performances to clarify the importance of metacognition in cooperation is compared.

A. Experiment

1) *Task*: An air traffic control simulator was used for a task (Fig. 2). The standard operating procedures of the task were as follows: (1) to select an aircraft with a mouse, and (2) to enter a command for the selected aircraft using a keyboard. Participants were asked to route arriving and departing aircraft both safely and accurately. During the session, the aircraft randomly appeared on the display. The participants were required to perform different sub-tasks simultaneously, such as understand commands that were given to aircraft, provide appropriate commands to control the altitudes and flight directions of aircraft, check distances between aircraft, make timely decisions about landings and takeoffs, monitor aircraft exiting from the airspace, etc. Each two-person team comprised a "Selector," who had only a mouse, and a "Commander," who had only a keyboard. The Selector selected the aircraft to which they would give a command with the mouse. Then, the Commander would enter a command for the selected aircraft using the keyboard. A team member could not complete these tasks by himself or herself and, thus, was required to cooperate with the partner. Because the number of aircraft increased in the second and third sets, team members had to reallocate team resources in the second and third sets; otherwise, they would fail to manage the aircraft.

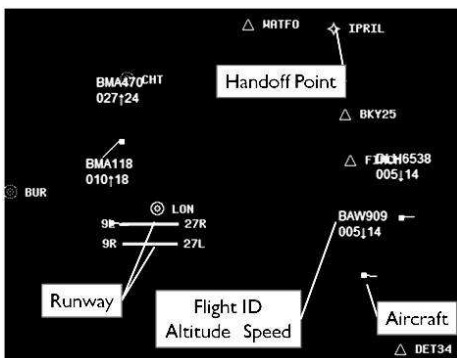


Figure 2. The simulator display

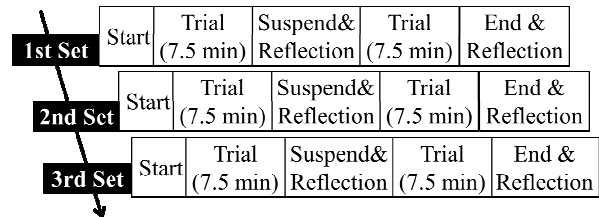
2) *Participants*: Twenty-six graduate/postgraduate students (13 teams) participated.

3) *Instructions for reflection*: Two-types of metacognitive instructions were designed and applied in order to investigate whether differences of metacognition in cooperation can affect team behaviors. The participants were asked to reflect on these instructions.

TABLE II. METACOGNITIVE INSTRUCTIONS

Instruction	Description
Team-oriented instruction	How is this task being operated by your team?
Self-oriented instruction	How do you cooperate in this task?

4) *Procedures*: The participants practiced the operation until they could smoothly land and transfer an aircraft. The total trial duration was 15 min for all participants. The metacognitive instruction was presented every 7.5 minutes and the participants read it and wrote down their own cognitive status and beliefs twice in each set (Fig. 3). When the instruction was presented, the display turned blank and the simulation was suspended. The participants sat face-to-face, and communicated freely with each other, except when they were responding to the instructions. Some teams could not participate in the third set because of their schedules.



**\*Reflection: the subjects read a metacognitive instruction and reflected upon their states of mind.**

Figure 3. Procedure

5) *Game scores*: Two types of game score were used as team performance indexes: safety violation time and number of aircraft successfully processed. Safety violation time was the duration in seconds of when the distance between two different aircraft was less than 1,000 feet vertically and 3 miles laterally. The number of aircraft successfully processed was calculated by subtracting the number of failed landings or improper exits from the airspace from the number of successful landings or successful transfers to other airspaces at the handoff points.

B. Data analysis

The reflection data were coded by a collaborator (who was unaware of the research purpose), based on the categories of the proposed model. Initially, the reflection data were divided into two categories: related to cooperation and not related to cooperation. Subsequently, each reflection related to cooperation was represented as a set of subcategories of the three primary viewpoints. A

subcategory “someone (self or partner)” was added in the “Subjects” category to code the reflection correctly. There were two types of sets of subcategories; a set of “Subject” and “Contents” and a set of “Subjects,” “Contents,” and “Metacognitive Skills”. For example, the reflection “(I am) monitoring what my partner is not monitoring” was coded as “Self + Cognition + Adjustment,” and the reflection “We demonstrate better performance than ever” was coded as “Team + Status + Evaluation (Good)”. In addition, the reflection data that were not related to cooperation were represented as “Self” and “Contents” or a set of “Self,” “Contents,” and “Metacognitive Skills”.

IV. RESULTS AND DISCUSSION

To investigate the importance of metacognition in cooperation, we conducted two types of analysis. In the first analysis, we compared both the reflection and the transitions of the game scores between the team-oriented instruction and the self-oriented instruction (Analysis I). In the second analysis, we compared the coded results of the reflection with the mean scores of each team to clarify the effect of metacognition on the cooperation of members with relatively good performance across the sets (Analysis II).

A. The Scores

In order to compare the transitions of the scores of the teams that obtained similar scores in the first set between the two metacognitive instructions, the teams were classified into three groups according to the number of aircraft successfully processed in the first set: high (6~10), middle (3~5), and low (0~2). Because the simulator was accidentally stopped during the experiment by one team for each instruction, their two values of performance data were excluded from the comparison on the scores. The teams that answered the team-oriented instruction and the self-oriented instruction were named T1~T6 and S1~S5, respectively.

Table III shows the transitions of the scores of the teams. T1 and S1 got the same degree of both the number of aircraft successfully processed and the safety violation time in the first set. Although T1 improved in both indexes in the second set, S1 deteriorated in the number of aircraft successfully processed. T2 and S2 got the same degree in both indexes in the first set. Both teams deteriorated in the safety violation time in the second set; the degree of deterioration in S2 was higher than that in T2. In the third set, both indexes of S2 improved. Both indexes of T3 and S3~S5 were similar in the first set; however, the safety violation time of T4 was longer than that of T3 and S3~S5. Although the performance indexes of both T3 and T4 improved in the second and third set, respectively, those of S3~S5 worsened in the second (S5) or third (S3 and S4) set.

In addition, in order to highlight the teams that got relatively good scores in the experiment, we divided the teams into two groups according to their mean scores in the experiment sets. Fig. 4 shows the mean scores of all sets for each team. T1~T4, S2, and S4 were categorized as teams that relatively performed well. On the other hand, T5, T6, S1, S3, and S5 were categorized as teams that relatively performed badly. Because the number of the scores was too small to

apply significance tests between the groups and between the sets, we could not discuss the significance of the differences.

TABLE III. THE SCORES

	Team	Scores	First	Second	Third	
High	T1	Success aircraft*	9	13	--	
		Safety violation time	647	189	--	
	T2	Success aircraft	6	6	--	
		Safety violation time	27	151	--	
	S1	Success aircraft	10	3	--	
		Safety violation time	779	192	--	
	S2	Success aircraft	7	7	10	
		Safety violation time	140	400	190	
Middle	T3	Success aircraft	3	7	8	
		Safety violation time	57	269	73	
	T4	Success aircraft	4	11	15	
		Safety violation time	475	435	414	
	S3	Success aircraft	4	8	-1	
		Safety violation time	140	400	190	
	S4	Success aircraft	5	9	3	
		Safety violation time	0	0	76	
	S5	Success aircraft	3	3	8	
		Safety violation time	27	879	170	
	Low	T5	Success aircraft	0	4	--
			Safety violation time	85	150	--
T6		Success aircraft	0	4	2	
		Safety violation time	293	138	125	

\* “Success aircraft” implies “number of aircraft successfully processed”.

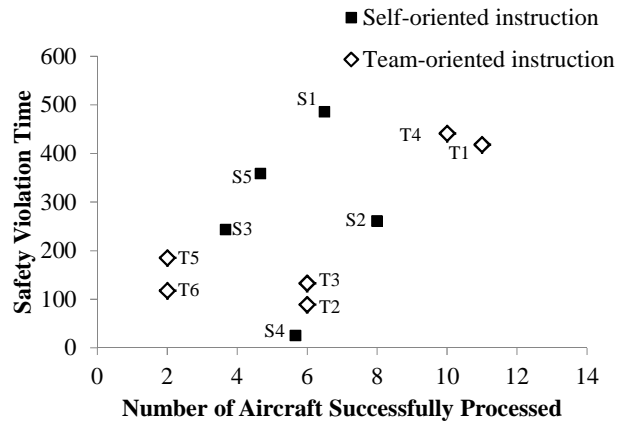


Figure 4. The mean scores

B. Analysis I

We statistically compared the two groups of the reflection to investigate the effects of the two types of metacognitive instructions on the viewpoint of metacognition pertaining to cooperation. Then, the game scores that were regarded as team performance indexes were compared.

Table IV shows the results of coding the reflection data by a collaborator who was unaware of the research purpose. We compared the mean number of each subcategory that was used to code the reflection elicited from the team-oriented instruction with that of the self-oriented one, using t-test per answer sheet for each member in each interval. Although the mean number of the reflection that does not relate to

cooperation in the self-oriented instruction was significantly higher than that in the team-oriented instruction, the mean number of the reflection that relates to cooperation in the self-oriented instruction was significantly lower than that in the team-oriented instruction. The mean numbers of “Partner,” “Team,” “Role,” and “Environment + Characteristics” in the team-oriented instruction were greater than those in the self-oriented instruction were. The coding pattern of “Role + Improvement” only existed when coding the reflection data that were derived from the team-oriented instruction. These results suggest that the team-oriented instruction induced metacognitive skills for cooperation more than the self-oriented instruction did.

Although the scores in the second and third set of the team-oriented instruction team improved, those of the self-oriented scores instruction team remained the same or worsened. The performance indexes after the team-oriented instruction may have improved because three teams (T2, T3, and T5) talked about their teamwork (e.g., role sharing) just after the team-oriented instruction. The team-oriented instruction induced team members to think about not only their own cooperation but also that of their partner and their team. This wide scope of metacognition in cooperation is expected to provoke metacognitive skills in both their

activities and interactions; thus, it probably encourages team members to find problems in their teamwork and strive to improve their team performance.

C. Analysis II

Table V shows the coded results of the reflection for each member. The number in each cell represents the sum of subcategories that were used for coding the reflection through all the sets. Because the differences between the metacognitive instructions affected the viewpoint of metacognition pertaining to cooperation, we compared the coded results of the reflection with the scores for each team within the teams that answered the same instruction.

Among the team-oriented instruction teams, T1~T4 were categorized as better-performing teams. On the other hand, T5 and T6 were categorized as teams that performed relatively badly. The fact that the reflections of T1 and T2 were coded using more “Self” and “Partner” values implies that the members of both T1 and T2 tended to reflect their cooperation in terms of the reductive perspective. Their reflection, which included “Self/Partner + Equipment + Characteristics,” “Compare,” and “Adjustment,” implied that the members applied metacognitive skills that were derived from the metacognition of their partner. Although the

TABLE IV. COMPARISON OF THE REFLECTION

	Team-oriented Instruction	Self-oriented Instruction	<i>p</i>
Not related to cooperation	0.98	1.88	t(100) = 2.60, <i>p</i> < .05
Related to cooperation	4.48	2.60	t(95) = 2.89, <i>p</i> < .01
Self	1.57	1.27	<i>ns</i>
Partner	1.04	0.46	t(78) = 2.43, <i>p</i> < .05
Team	1.68	0.88	t(98) = 2.67, <i>p</i> < .01
Role + Improvement	0.71	0.00	—
Environment + Characteristics	0.55	0.17	t(87) = 2.20, <i>p</i> < .05

TABLE V. THE CODED RESULTS OF THE REFLECTIONS

Team	Team-oriented instruction												Self-oriented instruction									
	T1		T2		T3		T4		T5		T6		S1		S2		S3		S4		S5	
Members	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S
Related to cooperation	18	26	20	33	14	20	18	2	36	21	10	16	14	19	35	10	29	17	26	7	3	3
+ Self	5	4	4	14	5	9	8	1	18	3	0	2	6	10	20	4	8	9	9	6	3	1
+ Partner	6	3	6	9	2	3	1	1	16	4	0	2	3	4	6	0	4	1	7	1	0	0
+ Team	3	13	10	10	5	8	9	0	2	14	10	12	5	5	9	6	17	7	10	0	0	2
+ Adjustment	0	1	0	0	0	0	1	0	0	0	1	1	0	0	5	0	1	0	2	1	0	0
+ Compare	0	3	2	2	0	1	0	0	0	0	0	0	1	0	2	0	4	0	0	0	0	0
+ Belief in Partner	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0
+ Improvement	0	0	2	2	4	4	2	0	0	5	2	0	4	2	1	1	0	1	3	1	0	0
+ Evaluation, Issue, Clear/Unclear	0	2	2	6	2	5	4	0	4	2	2	2	3	9	13	6	10	2	15	0	2	3
+ Self/Partner + Equipment + Characteristics	1	1	3	9	0	2	2	1	1	0	0	0	2	0	0	1	0	1	1	0	0	0
Not related to cooperation	0	1	4	0	18	2	10	22	3	9	0	0	8	12	7	25	3	6	3	4	10	6
+ Improvement	0	0	2	0	3	0	0	4	0	1	0	0	0	0	0	8	0	0	0	0	0	0
+ Evaluation, Issue, Clear/Unclear	0	1	1	0	3	1	10	6	1	3	0	0	6	8	5	13	0	2	2	0	3	1

\* “C” means “Commander” and “S” means “Selector”.

reflection of T5 included more “Self” and “Partner” values, the metacognitive skills of “Compare/Adjustment” were not included. The reflection of T5 included relatively fewer descriptions of cooperative relationships than those of “Self/Partner + Equipment + Characteristics,” and a variety of metacognitive skills including “Evaluation, Issue, Clear/Unclear,” and “Improvement”. The reflection of T3 included relatively more “Evaluation, Issue, Clear/Unclear,” and “Improvement” values. The number of reflections of T6 was relatively lower. Members of T6 tended to describe themselves in terms of “Team,” and there are few descriptions that included metacognitive skills. Among the self-oriented instruction teams, S2 and S4 were categorized as better-performing teams. On the other hand, S1, S3, and S5 were categorized as relatively bad. The reflection of S2 included more descriptions of metacognitive skills. Specifically, the commander’s description in S2 included a variety of metacognitive skills that were probably derived from belief in the partner, including “Adjustment,” “Compare,” and “Belief in Partner”. The reflection of S4, whose team performance indexes improved in the second set, but worsened in the third set, included descriptions of the characteristics of the equipment and “Adjustment”. Although the reflection of the commander in S4 included metacognitive skills through all the sets, the number of reflections that included metacognitive skills gradually decreased in the reflection of the selector. Although the reflection of S1 included more descriptions about cooperation and metacognitive skills such as “Compare,” there were no descriptions coded as “Adjustment” and “Belief in Partner”. There was a reflection coded as “Self/Partner + Equipment + Characteristics”; however, it was included by the final reflection. The reflection of S3 included more descriptions in terms of “Team”. Although the commander of S3 applied metacognitive skills in team cooperation, the selector did not. There were fewer reflections of S5.

These results imply the following four characteristics of the reflection in teams that had good cooperation: (1) members can easily understand the characteristics of their equipment; (2) members can describe their cooperation activities in terms of “Self” and “Partner,” rather than “Team”; (3) both members applied metacognitive skills in cooperation; and (4) metacognitive skills that can be derived from belief in partner, such as “Compare,” “Adjustment,” and “Belief in Partner,” can be important for positive teamwork. These four characteristics mentioned above may be compatible with past findings from research on teams: (1) can correspond to system monitoring, or can understand environmental characteristics in teams [9]; both (2) and (3) can correspond to mutual performance monitoring; and (4) can correspond to shared mental models and backup behaviors [10]. Observational marker systems for teamwork are probably not suited to evaluating these implicit teamwork behaviors. The reflection on cooperation and its analysis can be applied as a teamwork measurement for implicit teamwork and can be expected to give us good insights on some problems of team cooperation that cannot be identified through observation.

The team task used in the present study demanded team resource management, such as building effective cooperation patterns and adjusting behaviors, to not interfere with the partner, but to help the partner. It was necessary for the management to monitor the members’ status and to identify problems with their team cooperation. In addition, understanding the characteristics of members’ equipment in the task was necessary for building an effective cooperation style in this task. These are probably the reasons why teams whose reflections included more metacognitive skills and richer descriptions of cooperation could show and maintain good performances. If this is true, a metacognitive instruction that induces a wide range of metacognition in cooperation can be applied as effective team training in tasks that have the same characteristics as those in this experiment.

## V. CONCLUSION

This study aimed to examine the importance of metacognition in cooperation. The comparisons of both the reflection and the transition of the game scores between the team-oriented instruction and the self-oriented instruction suggested that a wide range of metacognition in cooperation could enhance team performance. In addition, the comparisons between the coded results of the reflection and the scores for each team suggested that the reflection of teams that had relatively good performances through the sets included metacognitive skills that were derived from belief in the partner’s cognition and activities. These two comparisons suggested that for positive teamwork, it is important that team members apply metacognitive skills to a variety of “Subjects” and “Contents” with each other.

## ACKNOWLEDGMENT

We express our deepest gratitude to the Japan Society for the Promotion of Science (JSPS) that supported this study.

## REFERENCES

- [1] A. Mishra, K. Catchpole, and P. McCulloch. “The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behavior in the operating theatre”. *Qual Saf Health Care*, vol. 18(2), pp. 104-108, 2009.
- [2] R. Flin and L. Martin. “Behavioural markers for crew resource management”. *Int. J. of Aviation Psychology*, vol. 11, pp. 95-118, 2001.
- [3] K. Nonose, T. Kanno, and K. Furuta, “A team cognition model derived from an analysis of reflection on cooperation,” *Cognition, Technology, and Work*, in press.
- [4] N.J. Cooke, J.C. Gorman, and J.L. Winner. “Team cognition,” in F. Durso, R. Nickerson, S. Dumais, S. Lewandowsky, and T. Perfect (eds) *Handbook of applied cognition*, 2nd edn. Wiley, London, pp. 239-268, 2007
- [5] T.O. Nelson and L. Narens, “Why investigate metacognition?” in *Metacognition: knowing about knowing*. J. Metcalfe, A. P. Shimamura (eds) MIT Press, Cambridge, MA, 1994, pp. 1-25
- [6] R. Tuomela and K. Miller, “We-Intentions”. *Philos Stud*, vol. 53, pp. 367-389, 1988.
- [7] W. McDougall, *The Group Mind*. Hew York; Putnam, 1921.
- [8] S.R. Chant and Z. Ernst Z, *Group intentions as equilibria*. *Philos Stud*, vol. 133(1):pp.95-109, 2007.



- [9] V. Rousseau, C. Aube, and A. Savoie, "Teamwork behaviors: A review and integration of frameworks". *Small Gr Res*, vol. 37(5), pp. 540–570, 2006.
- [10] E. Salas, D. E. Sims, and C.S. Burke, "Is there a "Big Five" in teamwork?" *Small Gr Res*, vol. 36(5), pp. 555-599, 2005.

## Educational Playability

### Analyzing Player Experiences in Educational Video Games

*Amer Ibrahim, Francisco Luis Gutiérrez Vela, José Luis González Sánchez, Natalia Padilla Zea*

Dept. of Software Engineering  
University of Granada  
Granada, Spain  
(ameribrahim, fgutierr, joseluisgs, npadilla)@ugr.es

**Abstract**—Educational video games constitute some of the main edutainment applications currently on the market. However, the development of video games as educational tools is very difficult due to their multidisciplinary nature (fun and education). Player Experience is a good measure of the level of fun and education presented to players and determines, to some extent, the success of an educational video game. This paper examines the importance of the role of Player Experience in motivating and enhancing the concept of learning through play, and highlights the role of Educational Playability attributes as suitable and effective tools to analyze and measure the experience obtained by a player during a game. The paper also discusses the role of a balanced Educational Video Game that provides a good level of motivation and engagement in order to improve player experience.

**Keywords**- *Playability; Player Experience; Educational Video Game.*

#### I. INTRODUCTION

Videogames are complex environments that offer the kind of entertainment and challenges that appeal to a growing number of people. The objective of videogames is to engage players in a fun activity which provides distraction from their daily lives, and which captures their attention for hours. We define Educational Video Games (EVG) as video games with special characteristics to teach, build and strengthen skills while the player has fun. Their focus on motivation, curiosity, emotion, socialization, entertainment and education distinguishes them from other types of video games. Play is fun, and information that is learned in an enjoyable way is less likely to be forgotten [23]. A very important aspect of the success of an EVG is how much it motivates players, as motivation guides learning activities during playing time. The inclusion of educational content often leads to a loss of the game's entertainment value, which decreases its efficiency as an educational tool [33].

In this paper, we examine the problem that many EVGs do not provide motivation to play, and thus do not improve the Player Experience PX (user experience in a video game). However, do all EVGs demonstrate this problem? Where exactly does the problem lie? How can we detect it? What does player motivation depend on? How can we determine

whether a game provides the necessary level of fun and motivation?

We believe this last question depends largely on the property of playability, which assesses and measures indirectly how much fun the game is and how much it attracts and captivates the player. To assess the level of playability, we must analyze the experience obtained by a player in an EVG. However, is the current structure of playability in video games sufficient to analyze the PX in EVGs? Is it necessary to qualify this structure? In this paper we answer these questions using a theoretical approach.

We aim to present educational playability (playability in EVGs) as a synthesis of all the playability studies analyzed throughout this work, and analyze the PX in order to obtain a measure of the fun and education experienced by a player with an EVG.

The subject of this paper overlaps many current topics in the games world, such as: User Experience, Video Game, Education and Human Computer Interaction.

Following this introduction, in Section 2, we discuss some studies of EVGs, present the characteristics of EVGs, and analyze the attributes of a balanced EVG. In Section 3 we present an analysis of the PX, a definition of Educational Playability, the attributes and facets which characterize and measure PX in EVGs, and define the correlation between playability attributes. Finally, we present our conclusions in Section 4, followed by our references.

#### II. EDUCATIONAL VIDEO GAMES

Many researchers have discussed the use of video games as useful educational tools to improve learning and performance. This is due to the many benefits video games have with regard to the attention and motivation of players, leading to greater academic success, better cognitive skills, and improved concentration. Entertainment is used in combination with education in order to create a motivating and successful environment for learning. [1] [24].

Gee [9] emphasizes the use of games as a medium to promote experiential learning. Gee discusses the opportunities that video games provide for simultaneous learning, by presenting three components for active learning: “*experiencing the world in new ways, forming new affiliations, and preparation for future learning*”. He asserts that video games teach very effectively through an environment that is demanding yet attractive, and that they

help keep a player positively engaged and motivated to continue learning. The concept of the EVG has a long history [19][34]; several attempts have been made to incorporate education into video games, such as Edutainment, Serious Games and Commercial off-the-shelf (COTS) [8][26], which effectively created games as a teaching tool specifically in the domain of computer science .

Nowadays, there are many EVGs. However, are all EVGs successful and do they always keep the player motivated to play? Good EVGs hook their users and maintain higher levels of motivation for longer. Robertson [25] highlights the value of educational games in promoting traditional literacy skills, including narrative structure and creativity, while providing significant benefits in terms of motivation, team work and self esteem. Moreover, the positive experience felt from self-initiated activities can lead to a greater sense of self-efficacy, which in turn increases a student's motivation to learn.

#### A. Characterization of Educational Video Games

EVGs are a combination of playing and learning, however, it can be difficult to give a precise definition of an EVG since there are many kinds, all very different from one another (simulator, social game, etc.). Not only is it necessary to identify the features that make EVGs different from other videogames, but a structure capable of achieving the objectives of EVGs is also required. There are several works in the literature that incorporate new features for building playful and attractive EVGs. Many researchers find that motivation to play is a very important feature of EVGs, based on game rules, goals, and rewards. They argue that competence, autonomy, and relatedness are factors that affect motivation [5]. Dicky and Waraich [6][30] mainly focus on the game narrative. They believe that a narrative context that promotes “*challenge, fantasy, and curiosity*” and provides feedback for players is one that promotes intrinsic motivation to play. Other works provide the social aspects of EVGs by developing skills and encouraging collaboration, cooperation and communication, so that players communicate with each other and work together to achieve common objectives. Bailey [2] suggests communication may be built or reinforced by video games through their capacity to offer a more social approach to learning and collaboration.

Malone and Lepper [19] identify four key characteristics which make an EVG enjoyable: challenge, curiosity, control and fantasy Thiagarajan [28] also presents the following characteristics: conflict, control, closure, contrivance and competency. Jenkins [14], identifies several characteristics of successful educational games such as: EVGs are made to fit specific learning contexts; successful EVGs supplement classroom teaching rather than replace it; using play as learning strategy; every element of the game design is meaningful; success is social, rather than individual; the game must be fun. Gee [10] proposes seven features of the videogame design that are important for effective learning: Interactivity, Customization, Strong identities, Well-sequenced problems, A pleasant level of frustration, A cycle of expertise, ‘Deep’ and ‘fair’.

Based on the above works, we here highlight the main characteristics that should be included in EVGs:

- Accelerate learning times and focus on reinforcement of players’ skills and experience.
- Increase player motivational levels by using accurate, richly detailed, predictable content.
- Increase players’ retention levels.
- Make players feel that they are part of a creative and dynamic community.
- Give players great scope for personalization according to learning style and preferences.
- Present suitable and effective content for players of all levels.
- EVG is an active virtual world, which provides an active support system and feedback and reacts to players’ actions in a consistent, immediate, challenging and exciting way.
- Use collaborative learning and improve collaboration and interactivity among players.
- Transmit different stimuli to activate and engage players during playing time.

These characteristics represent the important role that EVGs play as interactive tools to motivate, teach and entertain players. Games that do not captivate and engage players will not motivate them to play, meaning that their time will not be utilized effectively.

#### B. Balanced EVGs

Balancing fun and engagement with educational content and learning is one of the most important challenges that EVG designers face. “*We want the game, as much as possible, to be time spent learning*”, “*You have to take both engagement and learning into account*” [36]. Thus, an EVG is controlled by its success as a tool to teach and entertain. The success of an EVG is related to a number of factors that balance the playful content and educational content.

One kind of content should not be given more importance than the other, since both form the structure of the EVG and ensure its success or failure. Law [17] indicates that one of the problems of current EVGs is the poor balance between playful and educational activities, or between challenges and ability. Moreover, the lack of sound instructional models, based on pedagogical standards and didactical methods, is seen as the common weakness of most EVGs, leading to a separation of learning from playing. The imbalance between EVG components may lead to the failure of an EVG to achieve its objectives. Other reasons for the failure of EVGs can be identified as: uninteresting content, the fact that presented contents are not appropriate for the prior skill and knowledge level of players, and game elements and game contents are unrelated.

In this work, present new factors which characterize the relationship between EVG components, based on the characteristics previously mentioned:

1) *Interdependency*: The success of an EVG depends on the relationship between educational content and playful content, and how these contents are presented without either one dominating at different stages It also depends on

keeping players interested and motivated to play. Morenoger notes that the different designs found in this field have aims that seem to be biased towards either fun or educational content [21]. The proposed content must support the achievement of the game objectives, which means that EVG success is constrained by the interdependence and integrity between playful and educational components during all game steps, as well as the continuity of the game story and the compliance of game contents in terms of the visualization and achievement of goals. Prensky states that effective educational game design must achieve a balance between fun and educational values [23].

2) *Continuity and consistency*: Keeping the Learning-Fun Relationship balanced during all game steps in a way that engages players. EVGs that are not engaging can negatively affect players as their time will not be utilized effectively. We therefore emphasize that an EVG must contain different goals and objectives (playful and educational) throughout the different levels, and well-structured and carefully-chosen content during the design step of the game environment. It is necessary to provide appropriate content for the predetermined learning objectives and player skills. *“The best way to learn is when the learner is having fun at the same time. Having fun gives your kid motivation to keep on practicing, which is the only way to learn skills”* [37].

3) *Completely New*: In addition to the previous factors, the EVG content should be renewed each time the player plays (i.e., use different ways to present the EVG contents), the game must keep the player immersed and provide new knowledge, experience, etc. This, along with a good, realistic presentation of game content, means that the player loses his or her sense of time and place. Mavis Beacon [38] provides typing tests with different kinds of presentation (graphics, sounds, etc.) throughout the game in order to capture a player’s attention and encourage him or her to pass the tests.

These factors indicate that to develop a playable EVG and improve the overall PX it is important to understand the changes that have occurred in the video game structure due to the merging of educational and fun components. Due to its effect on the learning progress, we need to characterize PX in EVGs in order to know which elements of the video game are helping or hindering its success.

### III. PLAYABILITY: A MEASURE OF A PLAYER EXPERIENCE IN EDUCATIONAL VIDEO GAMES

Above, we have mentioned the different aspects of EVGs (hedonics, motivation and the narrative structure) which help to improve PX by teaching higher order thinking skills such as strategic thinking, interpretative analysis, problem solving, plan formulation and execution, adaptation to rapid change, and offer opportunities for players to cooperate within the game and within the larger community through websites and online forums [26].

However, measuring experiential dimensions such as fun, motivation and emotion is more elusive than measuring traditional performance metrics such as time spent playing,

or number of tasks completed in each step. “It is difficult to obtain knowledge about what players did when playing the game, and how meeting different game design elements affected their experience of interacting with the game” [29]. Thus, to measure the PX, we propose the use of playability attributes and properties that allow us to measure whether a player is having fun playing an EVG based on the pragmatic and hedonic properties. Playability attributes characterize PX by analyzing all aspects of player behavior and feelings when he or she is playing. An adequate value of playability allows the player to have a more positive game experience, resulting in a greater predisposition to assimilate the educational and playful concepts underlying the game dynamics. Accordingly, we need to define the PX and identify the playability attributes and properties in the EVG context to ensure an optimal PX.

The Player Experience PX is a simulation of User Experience UX in software application used to perform daily tasks. UX is defined as how a person feels when interacting with a system in terms of ease of use, perception of the value of the system, utility, and efficiency in performing tasks etc [27]. PX may be much more extensive than the UX; PX helps to improve the player interface design, the level of challenges, the pace, game mechanics and game story. We can say that PX extends the UX in EVGs, where UX is focused more on the subjective part of the interaction process and goes beyond the traditional study of skills and cognitive processes of users and their rational behavior when interacting with computers [18].

The term PX, based on definitions of user experience [18], refers to “all aspects related to the player that are affected by and interact with the playing environment”. These aspects represent the different features of the interaction process, such as: sensation, feelings, emotional response, assessment, user satisfaction and the experience obtained throughout the entire playing time [13]. A player experiences these aspects positively in a videogame with good playability, where the playability attributes and properties are measures of PX. Canossa [3] emphasizes the emotion and narrative aspects of the game in the design of a UX. Canossa describes how specific game environments and certain game features can offer players personal styles of play. To analyze the quality of a video game we must discuss the non-functional values, which define specific properties of game and affect its quality [12]. Playability has been studied from different perspectives and with different objectives without consensus on its definition or the elements that characterize it. In this context, Playability is a term used in the design and analysis of video games that describes the quality of a video game in terms of its rules, mechanics, goals, and design. It refers to all experiences that a player may feel when interacting with a video game system. Playability is affected by the quality of the storyline, responsiveness, pace, control, intensity of interaction, intricacy, and strategy, as well as the degree of realism and the quality of graphics and sound [32][35].

There are few studies focused on defining Playability formally, the exceptions being the work of Fabricator and Järvien[7][15]. However, these do not specifically refer to

Playability attributes or the properties which characterize it. Järvinen has defined playability as an evaluation tool which consists of four components: functional, structural, audiovisual, and social playability [15]. Fabricatore has defined playability in action games as the possibility of understanding and controlling gameplay [7].

Playability has been defined as “*the instantiation of the general concept of usability when applied to videogames and it is determined by the possibility of understanding or controlling the gameplay*” [7]. This may be interpreted as supporting a separation of playability and usability if we consider that playability is related only to video games, “*capability to provide enjoyment for a player over an extended period of time*” [16]. Thus Playability is based on Usability but, in the context of video games, it goes much further. Furthermore, Playability in video games is not limited to the degree of fun or entertainment experienced when playing a game; it also extends the UX characteristics with player dimensions using a broad set of characteristics such as motivation, pleasure, curiosity, emotion, and social influences that contribute significantly to user satisfaction with the product

#### A. Educational Playability

The main concepts studied in works that focus on playability experiences are: flow [4], immersion [22], emotions [20], and fun [31]. Are these aspects sufficient to characterize playability in EVGs? EVGs are composed of two dimensions: educational and playful. Thus, Playability in EVGs is not limited to playful objectives but must take into account educational objectives, such as learning while having fun, improving the abilities of students to solve complex problems, reinforcing players' skills and improving player experience.

EVG must maintain fun as a subjective concept and include educational goals. Thus, Educational Playability in EVGs is an extension of playability in videogames. We propose the definition of playability in EVGs to be “*the set of properties that describe the PX in the gaming environment, whose main goal is to provide fun and learning in a playable and learnable context during the entire playing time*”. Educational Playability describes how to make an EVG playable and learnable at the same time, and it includes two aspects: learning to play and playing to learn. Thus, it contains new characteristics related to the educational component.

Characterizing Educational Playability (i.e., the playability of the playful and educational contents in EVGs) requires a specific study related to the different aspects of an EVG (analyzing several EVGs and their characteristics). Moreover, characterizing Educational Playability can result in improved player motivation and satisfaction, the establishment of rules of engagement, the introduction of educational content that is implicit and relevant to the narrative story line, improved quality of educational content, the provision of interactive feedback and an emphasis on the correctness, accuracy and reality of educational content.

To characterize Educational Playability we have used a tested and verified characterization proposed by González

Sánchez [11][12] to achieve the new changes in the playability structure. We have therefore identified and qualified new attributes and properties which are suitable for defining playability in EVGs.

Fig. 1 shows the educational playability attributes divided into educational attributes (the newly proposed attributes) and playful attributes.

#### B. Educational Playability Attributes

We will now discuss in more detail the new playability attributes for characterizing the educational component of PX (educational attributes), taking into account the educational aspects proposed in this paper. We have already mentioned the effects of the educational content on all other playability attributes (playful attributes), and have presented some of the possible ways that each attribute can be measured. Thus, Educational Playability is characterized by:

1) *Satisfaction*: The gratification or pleasure derived from playing a complete video game or some aspect of it. Satisfaction is an attribute with a high degree of subjectivity. Not only is it difficult to measure, but it also influences player preferences. Satisfaction is linked to the degree of balance and harmony in the game. This is defined by the following properties: Game Fun, Contents Disappointment and Attractiveness. A high degree of each one of these properties will lead to an increase in player satisfaction.

We can measure this attribute by the percentage of game objectives discovered and achieved during the playing time and by the percentage of useful playing time (amount of time a player feels pleasure while playing).

2) *Learnability*: The player's capacity to understand and master the game system and mechanics (objectives, rules, how to interact with the video game, etc). It has the following properties: Game Knowledge, Player Skill, Game Difficulty, Player Frustration, Speed of Learning and Discovery Techniques. A high level of player knowledge and skills will reduce the degree of learning, and a high level of difficulty and frustration will directly decrease the playability of the game. However, rapid presentation of game concepts and the ability to discover these concepts will positively affect the degree of learning. This attribute introduces learning to play, but playing to learn is also a very important characteristic of EVGs. EVGs involve a complex learning process, and the importance of this means that we have added a new attribute related to educational aspects of the video game.

To measure Learnability we can use player performance and the playing time, which are related to mastering the game controls and mechanics.

3) *Effectiveness*: The resources necessary to offer players a new experience (fun and learning) while they achieve the game's various objectives and reach the final goal. To ensure enjoyment and hide the educational component in EVGs in a homogeneous way, we must include elements from a playful point of view. This will reduce the effectiveness of the educational process and it may seem that time is being wasted in playing, but we are in fact increasing the motivation to learn. Effectiveness has the

following properties: Game Completion and the Structuring of Game Resources. An EVG is more effective if the percentage of Completion and Structuring is high.

To measure this attribute we can use the percentage of unachieved objectives, and also analyze the gap between the prior knowledge of the player and the knowledge presented by the challenges in the game.

4) *Immersion*: The capacity of the EVG contents to be believable, such that the player becomes directly involved in the virtual game world. At an educational level, this property is used to measure the ability of an EVG to present the educational aspects implicitly. It has the following properties: Conscious Awareness, Absorption in game, Game Realism, Control Dexterity and Socio-Cultural Proximity with the game. We can keep a player immersed while playing by focusing on a high degree of these properties.

This attribute can be measured by the playing time, time per challenges, the level of player awareness, and the quality of the actions performed.

5) *Motivation*: The set of game characteristics that prompt a player to perform specific actions and continue undertaking them until they are completed. At an educational level, motivation to play indirectly produces positive motivation to learn. It can be defined by: Encouragement Techniques, Curiosity about the game, Player Self-Improvement and Diversity of game resources. These properties evaluate the player's perseverance to overcome the game challenges.

Motivation can be measured using the educational level obtained during the game, the time and number of attempts a player makes in order to overcome a game challenge, and the objectives achieved.

6) *Emotion*: This refers to the player's involuntary impulses in response to the EVG stimuli that induce feelings or a chain reaction of automatic behaviors. The educational content in EVGs may provoke rejection by the player, which reduces the motivation for the player to explore the game and thus achieve the educational goals. Emotion contains the following properties: Player Reaction, Game Conduct and Sensory Appeal for game elements. These properties are able to generate a set of emotions in the player during the playing time which are very difficult to achieve in the real world.

We can measure this attribute by analyzing the pleasure gained, and by studying the emotional reaction and educational development of the player when facing and overcoming the different challenges.

7) *Socialization*: The set of game attributes, elements and resources that promote the social dimension of the game experience in a group scenario. From an educational

perspective, socialization is the ability to support students learning from one another. Socialization relates to the following properties: Social Perception, Group Awareness, Personal Implication, the Sharing of Social Resources, Communication Techniques and Interaction Rules of the game. We can use these properties to present new challenges that help players become immersed in the virtual game world and feel satisfied with the game dynamic, and that help create collective emotions among players. Thus each player is self-stimulated and is encouraged by other players to face and overcome the proposed game challenges.

The Socialization attribute can be measured through an analysis of the social elements, rules, effects, mechanisms, and the percentage of objectives that have been achieved in a group situation.

The preceding attributes of Educational Playability are used to measure playful aspects, although they have been adapted to the special case of EVGs. The following two attributes are used specifically for measuring aspects of learning and its relationship to the playing process.

8) *Supportability*: we define this as the ability of EVGs to engage and teach players correctly, and encourage them to continue learning and achieve the learning objectives causing playability as motivational element. We propose the following properties to characterize Supportability:

a) *Embedded learning*: This is the capacity of the game to implicitly introduce the educational content and encourage players to learn it while playing. This property aims to ensure that the level of pleasure and fun is unaffected by the educational content.

b) *Interdependence*: This refers to the relation (interdependence) between educational content and the game elements, and is related to game activities, challenges and game objects that represent the objectives presented to a player during the game.

c) *Integrity*: The game must ensure that the relationship between playful and educational content is well managed, so that new elements presented for both contents are compatible with each other. Integrity gives players an idea of the learning process and how they can achieve the presented game content.

d) *Challenges*: Refers to the different levels of difficulty and complexity of challenges, and how to suit these to learners' skills, as well as how to relate these challenges to the presented content, so that flow experience is possible. A game with a high degree of complexity may cause the player to lose interest.

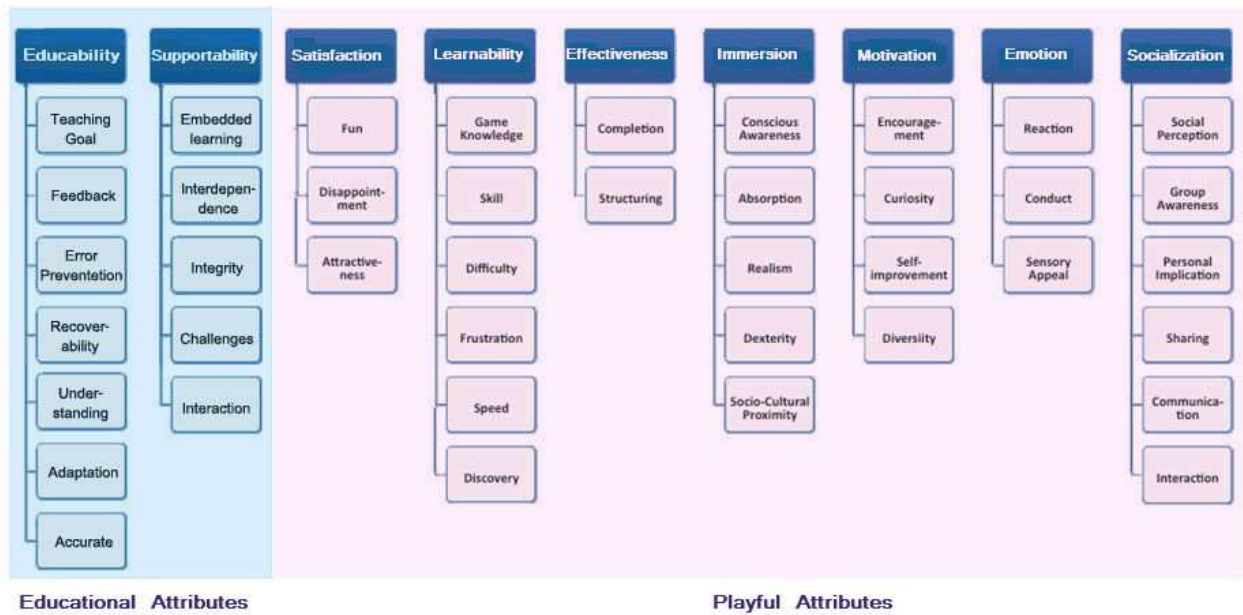


Figure 1. Educational playability Attributes.

e) *Interaction*: This refers to the ability of the game to increase players’ motivation to learn by encouraging them to interact with the educational content through quizzes and multimedia, stimulating them to practice skills, demonstrate their knowledge, discover information, and reinforce learning.

This attribute can be measured by determining the development of player knowledge and the percentage of educational elements provided in the game, and by analyzing the percentage of educational goals achieved.

9) *Educability*: We define this attribute as: the educational characteristics of video games that support the user’s ability to be aware, understand and master learning goals. We propose the following properties to characterize Educability:

a) *Learning goal*: the degree to which the educational content teaches and increases player skills, and the ability to use these skills. This property must motivate the learner to achieve the goals and learning experiences. Thus, an EVG that does not improve player knowledge does not teach players.

b) *Feedback*: Games must present appropriate educative feedback that improves the behavior, experience, and performance of players based on their progress and whether their decisions have positive or negative effects. This property aims to help players to be aware of and to achieve the educational content in the shortest possible time.

c) *Error prevention*: EVGs should ensure that a user’s skills develop correctly, never allowing a player to use incorrect information during the learning process or when

trying to solve tasks. Error prevention is defined by the degree to which the educational elements help players avoid making errors and achieve the content correctly. This increases a player’s efficiency and increases the achievement of educational goals. It also reduces the level of frustration and encourages players to play by making the game more acceptable to them.

d) *Recoverability*: The degree to which the EVG allows players to check the correctness of the content provided knowledge and change incorrectly and unintentionally acquired knowledge and skills. Players frequently choose incorrect options or actions and are likely to find themselves with incorrect knowledge in an error state from which they must recover.

e) *Understanding*: The degree to which the player is able to understand the game goals, determine the educational content in the game environment during playing time, and understand the difficulties at each game level. *Understanding* helps to define the user profile and determine the required feedback.

f) *Adaptation*: Games must be able to bridge the gap between a player’s prior knowledge and that which is presented at different game levels, encourage the player to use this knowledge to pass each level, and enhance the player experience in order to achieve the EVG’s objectives. the improvement of user experience and the success a player has at each game level will enable the player to more effectively achieve the target goals.

g) *Accuracy*: The degree to which correct and real content helps players to focus on the content provided and obtain new knowledge from the virtual world.

To measure this attribute we can use the amount of assistance (e.g., hints and tutorials) provided throughout the game and analyze player actions, performance and skills based on the knowledge gained during the game.

### C. Correlation among Playability Attributes

In this section we examine the correlation among educational playability attributes. For example, if an attribute has a positive value, it will generate a positive experience in all other attributes. This is due to the fact that the attributes are focused within the experience. Thus, the model we present characterizes the player's experience. By studying and analyzing each attribute, the relationship between attributes and the effects that attributes have on one another can be clarified.

1) *Satisfaction*: This is an indicator of the improvement in player skills and knowledge, indicating the degree to which a player is immersed in the learning and playing process. Satisfaction is related to how well game convinces the player to learn the content presented during the game, and his or her ability to control and master this content. If a player has high level of satisfaction this indicates that he or she does not find the educational content difficult to achieve.

2) *Learnability*: A low level of learnability means the player finds the game mechanics and game rules easy master. The player therefore plays with high level of satisfaction and is motivated to overcome the game challenges, discover the game objectives and interact with educational content.

3) *Effectiveness*: An EVG has a high level of effectiveness if the player has discovered and achieved a high number of playful and educational goals, which means the player is satisfied playing a game. It also indicates that the game has provided the learning mechanism correctly in the virtual game world, meaning that a player requires only a low level of educational support to be able to master the dynamics of the learning process.

4) *Immersion*: An EVG with a high degree of immersion means that the player is easily captivated by the virtual game world, increasing the success of the game and enhancing a player's ability to discover and achieve the educational goals. It also stimulates players to interact with, and thus understand, most of the game content.

5) *Motivation*: A high level of motivation encourages players to discover new objectives presented in the game. This has an important effect on enriching a player's skills and knowledge, and indicates how well the game manages the relationship between different elements of the game. A high level of motivation also refers to effective mechanisms for supporting players to face and overcome game challenges, and helps to introduce the educational content during the game without players becoming frustrated (embedded learning).

6) *Emotion*: A high level of emotion provides players with different sensations (e.g., pleasure, boredom, frustration, challenge) that lead them to interact with different game elements. This interaction keeps maintains a player's motivation to achieve the game objectives. If an EVG has a high level of emotion the player feels more immersed in the virtual world, which helps to introduce the educational content implicitly and provides greater encouragement, motivating the player to overcome a particular educational challenge.

7) *Socialization*: An EVG with a high level of socialization in a multiplayer environment contains many different educational playability attributes such as: increase the player's ability to achieve the game goals, increase motivation to play and interact with different game elements, help to discover the educational content implicitly presented, reduce the need for game feedback, decrease the probability of recoverability, and facilitate a player's mastering of the learning process.

8) *Supportability*: A high level of supportability means that the EVG has managed the balance between educational elements and playful elements perfectly. These elements transmit different types of emotions that guide player interaction with the game. It also encourages players to achieve the educational objectives by motivating them to perform specific actions in the virtual game world and helping them to complete as much of the game as possible. High supportability involves a high level of socialization, because an EVG increases collaboration and interaction among players when it helps them to achieve the game goals.

9) *Educability*: A low level of Educability relates to minimal effort required by players in order to understand the educational content. This means that the game controls, mechanics and rules are clear enough that the player can master them easily, thus ensuring a high level of satisfaction. A low level of Educability helps players to discover the different game resources, and thus increases the effectiveness of the game. It also helps players to control all characteristics of the virtual world, increasing player immersion and motivation. This is due to the fact that players do not need to learn how to play and thus have more confidence to face game challenges. The level of Socialization will be high if the required level of Educability is low in a multiplayer game, because the player does not need to learn the dynamics of a multiplayer game.

### D. Revolution, A Case Study of Educational Attributes

In this section our intention is to analyze the video game on American history "Revolution" [39], by using the educational playability attributes proposed in this paper. However, the limitation of this work means we will analyze the game at the educational level using only the educational attributes (Supportability, Educability) "Fig.1".

Revolution is a multi-player game about the American Revolution. The game was designed to teach 18th century social and historical events of the USA, with players taking roles as farmers, slaves, politicians or merchants living in the colonial town of Williamsburg.



1) *Supportability*: “Revolution” the American history is one of the most captivating educational video games currently on the market. At each level the game presents wonderful graphics, music and narration. The game storyline draws students in, and players do not consciously realize that they are learning a great deal about American history and what colonial life was like in the 1700s.

Revolution uses the relationship between EVG contents to captivate players. As previously mentioned, at each level it presents different elements of the playful content (narrative, graphics, sound and visual effects) along with the historical content (Educational Content).

There is a high level of interaction in “Revolution”, whereby every interaction with a character includes an impressive graphical representation of that person in cartoon form. Players have the opportunity to interact with characters by choosing responses at the bottom of the screen.

2) *Educational*: “Revolution” incorporates the historical content into the game play in an embedded way in order to encourage better retention of the information. Whenever players click on an object, they are provided with important historical information and facts.

“Revolution” includes important historic terms representing facts that students should learn about American history. The game encourages students to click on these terms, read the definitions and “collect” these “smart words”. In this way, it ensures the quality of the learning process. “Revolution” helps players to master and understand the historical content by producing a catalog of accessible, fun to play titles that give players some sense of the actual history of a given period, not the myths that have developed around them.

Revolution provides players with an experience of history that includes passionate rhetoric and heroic battle, but also economic frustration, political indifference, and the mundane affairs of everyday life.

Revolution does not directly support Recoverability and error prevention, but does so indirectly through feedback and the information presented to players during the game. Revolution does not adapt the presented content to a player’s skills; it presents a series of easily understood actions and events relating to the social, economic, and political aspects of daily life.

#### E. Educational Playability Facets

Educational video game and playability analysis is complex enough to warrant being examined from different perspectives. Facets of Playability in video games have been proposed in order to facilitate the analyzing process [11]. Each facet allows the identification of the different attributes and properties of playability that are affected by the different elements of the video game architecture. This means it is possible to manage the process of analysis by taking into account the different video game elements presented in the game.

Playability facets offer different ways of analyzing the playability among the various elements of an EVG. They offer the possibility of interactive level analysis (user

interface, menus, and controls), intrinsic analysis (rules, goals, challenges, rewards) or hedonic analysis (emotional, cultural factors). Furthermore, the facets act as a logical subdivision of global playability into more specific playabilities, which together identify the global playability of a videogame. The proposed set of facets is related to video game elements (game core, game engine, and game interface), and ensures the following video game aspects: aesthetic, interactivity, and social.

If we consider playability at the educational facet of an EVG, our analysis will show the main purpose of the game to be educational, with all incorporated elements designed to promote learning. Grouping all educational aspects logically in one facet helps to facilitate the analysis of the educational values presented, and thus facilitate the analysis and design of EVGs.

The overall Educational Playability of a game comes from the value of each attribute in the different playability facets presented. It should be adequate enough that a player’s experiences and feelings when playing are the best possible and the most appropriate to the nature of the game. It is therefore necessary to qualify the facets so that they include the educational attributes of playability affecting the educational component of the PX. The Facets of Educational Playability are:

1) *Intrinsic Playability*: This is the playability inherent in the nature of the videogame itself and how it is presented to the player. It is related to the rules, objectives, pace and game mechanics.

2) *Mechanical Playability*: This is related to the quality of the videogame as a software system. This facet emphasizes features such as the fluidity of the scenes, the correct lighting, sound, graphics, movements, behavior of the game characters and environment.

3) *Interactive Playability*: We distinguish two types of interactive playability: educative and playful. Playful is associated with the user interface design, the mechanisms of content presentation and control systems. Educative is related to the supporting mechanism and managing the presentation of educational content.

4) *Artistic Playability*: This facet relates to the quality of the artistic and aesthetic rendering of the game elements. It is related to the graphic and visual quality, sound effects, story and narrative form and the environment created by all these elements within the game.

5) *Educative playability*: This is related to the quality of the educational content that is presented in the game and how to execute it. This facet relates to educational content properties such as correctness and suitability to the EVG objectives, player profile, game reality, and accuracy.

6) *Intrapersonal Playability or Personal Playability*: This refers to the individual outlook, perceptions and feelings that the videogame produces in each player and as such has a high subjective value.

7) *Interpersonal Playability or Social Playability*: This refers to the feelings and perceptions of users, and the group

awareness that arises when a game is played in a group context, be it in a competitive, cooperative or collaborative way.

#### IV. CONCLUSION AND FUTURE WORK

EVGs are a pervasive cultural phenomenon. The inclusion of computer games within formal learning is a necessity and should be commonplace.

In this paper, we have analyzed the characteristics that make an EVG useful, in order to analyze its level of effectiveness. We have also demonstrated how we believe the success of an EVG can be assured, based on the Balance, Player Experience and Educational Playability, and drawing on our experience in this field to discuss these three points. We have presented the balance between the educational elements and fun elements as one of the most important aspects of an EVG to affect the PX, and have analyzed what this balance should be. We have also examined the vital role that PX has in video game development, discussing its importance in building successful and effective EVGs. To measure the PX we have used Educational Playability attributes and properties (playability in EVGs). Accordingly, we have defined a set of helpful attributes and properties of Educational Playability that can be used to better understand and analyze the EVG structure.

To facilitate the analysis of Playability, and to guarantee a high level of quality in Educational Playability and improve the PX we have intuitively and explicitly identified the Facets of Playability. These facets can serve as basis for understanding and designing new and better PX in video games. They are also helpful in identifying the elements needed to achieve good overall Educational Playability in EVGs by studying every property in each attribute.

We are currently working to present a catalogue of Design Patterns that includes as many common EVG problems as possible, and propose a new way of benchmarking the efficiency of design patterns by analyzing the role of these patterns to improve and strengthen the PX. We are also working to propose mechanisms for evaluating the PX in EVGs based on the evaluation of Playability. Accordingly, we will be working on specific heuristics and evaluation tools for the field of EVGs.

#### ACKNOWLEDGMENTS

This work is financed by the Ministry of Education and Science, Spain, as part of DESACO Project (TIN2008-06596-C02-2) and the F.P.U Program

#### REFERENCES

- [1] C. Aldrich, *Simulations and the future of learning. An innovative and perhaps revolutionary approach to e-learning.* San Francisco, Pfeiffer, 2004.
- [2] C. Bailey, E. Pearson, S. Gkatzidou, and S. Green, "Using Video Games to Develop Social, Collaborative and Communication Skills," Proc. World Conference on Educational Multimedia, Hypermedia and Telecommunications, Chesapeake, VA: AACE, 2006, pp.1154-1161.
- [3] A. Canossa, and A. Drachen, "Play-Personas: Behaviours and Belief Systems in User-Centred Game Design," proc. of Human-Computer Interaction INTERACT, Springer, 2009, pp.510-523, DOI: 10.1007/978-3-642-03658-3\_55.
- [4] B. Cowley, D. Charles, M. Black, and R. Hickey, "Toward an understanding of flow in video games". *Computers in Entertainment*, vol 6, 2, 2008, doi:10.1145/1371216.1371223.
- [5] G. Denis, and P. Jouvelot, "Motivation-driven educational game design: applying best practices to music education," Proc. ACM SIGCHI International Conference on Advances in computer entertainment technolog. ACM Press, 2005
- [6] M.D. Dickey, "Ninja Looting for instructional design: The design challenges of creating a game based learning environment," Proc SIGGRAPH'06, ACM, 2005.
- [7] C. Fabricatore, M. Nussbaum, and R. Rosas, "Playability in Video Games: A Qualitative Design Model," HUMAN-COMPUTER INTERACTION, vol 17, 2002, pp. 311-368.
- [8] K. Facer, T. Rudd, R. Sandford, and M. Ulicsak, "Teaching with games: Using commercial off-the-shelf games in formal education," 2006, Bristol: Futurelab, doi: [http://www2.futurelab.org.uk/resources/documents/project\\_reports/teaching\\_with\\_games/TWG\\_report.pdf](http://www2.futurelab.org.uk/resources/documents/project_reports/teaching_with_games/TWG_report.pdf).
- [9] J.P. Gee, "What video games have to teach us about learning and literacy," *ACM Computers in Entertainment*, vol 1,1, 2003, pp. 20-20.
- [10] J.P. Gee, Are Video Games Good for Learning? Keynote address at Curriculum Corporation 13th National Conference, 2006.
- [11] J.L. González Sánchez, *Playability and Video Games. Analysis and Design of User Experience on Interactive and Entertainment Systems (Jugabilidad y Videojuegos. Análisis y Diseño de la Experiencia del Jugador en Sistemas Interactivos de Ocio Electrónico)*, Ed. Académica Española, Lambert Academic Publishing GmbH & Co KG, 2011.
- [12] J.L. González Sánchez, N. Padilla Zea, and F.L. Gutiérrez Vela, "From Usability to Playability: Introduction to Player-Centred Video Game Development Process," Proc Human-Computer Interaction International (HCI), Springer, 2009, pp.65-74, DOI: 10.1007/978-3-642-02806-9\_9
- [13] J.L. González Sánchez, R.M. Gil Iranzo, and F.L. Gutiérrez Vela, "Enriching Evaluation in Video Games", proc INTERACT 2011, Springer, 2011, pp.519-522, doi: 10.1007/978-3-642-23768-3\_72.
- [14] H. Jenkins, From serious games to serious gaming. Message posted to Confessions of an Aca-Fan, 2007, doi: [http://henryjenkins.org/2007/11/from\\_serious\\_games\\_to\\_serious\\_1.html](http://henryjenkins.org/2007/11/from_serious_games_to_serious_1.html).
- [15] A. Järvinen, S. Heliö, and F. Mäyrä, *Communication and Community in Digital Entertainment Services. Prestudy Research Report. Evaluation*, University of Tampere Hypermedia Laboratory, 2002.
- [16] J. Kücklich, "Play and Playability as Key Concepts in New Media Studies," Citeseer, 2004, pp. 1-44, doi:10.1.1.94.4169&rep=rep1&type=pdf.
- [17] E. Law, M. Kickmeier-Rust, D. Albert, and A. Holzinger, "Challenges in the development and evaluation of immersive digital educational games," Proc 4th Symposium of the Workgroup Human-Computer Interaction and Usability, Springer, 2009, pp. 19-30.
- [18] E.Law, V. Rohto, M. Hassenzahl, A. Vermeeren, and J. Kort, "Understanding, Scoping and Defining User Experience: A Survey Approach," Proc Human Factors in Computing Systems CHI'09, ACM, 2008.
- [19] T.W. Malone, and M.R. Lepper, "Making learning fun: A taxonomy of intrinsic motivations for learning," In *Aptitude, Learning and Instruction III: Cognitive and Affective Process*

- Analyses, R.E. Snow and M.J. Farr, Eds. Hillsdale, N.J.: Erlbaum, 1987.
- [20] R.L. Mandryk, M.S. Atkins, and K.M. Inkpen, "A continuous and objective evaluation of emotional experience with interactive play environments," Proc SIGCHI conference on Human Factors in computing systems, ACM, 2006, pp.1027-1036.
- [21] P. Morenoger, D. Burgos, I. Martínez-Ortiz, J.L. Sierra and B.F. Manjón, "Educational game design for online education," Computers in Human Behavior, vol 24, 6, 2008, pp. 2530-2540., doi:10.1016/j.chb.2008.03.012
- [22] L. Nacke, and C.A. Lindley, "Flow and immersion in firstperson shooters: measuring the player's gameplay experience," proc Conference on Future Play: Research, Play, Share, ACM, 2008, pp. 81-88.
- [23] M.Prensky, Digital Game-based Learning (1 ed.). Two Penn Plaza, New York, NY 10121: McGraw Hill, 2001.
- [24] M. Prensky, Digital game-based learning, 2003, ACM Portal, 1(1).
- [25] J. Robertson and C. Howells, "Computer Game Design: Opportunities for Successful Learning," Computers & Education., vol 50, 2, 2008, pp.559 - 578, doi:10.1016/j.compedu.2007.09.020
- [26] D. W. Shaffer, K. Squire, R. Halverson, and J. P. Gee, Video games and the future of learning. Phi Delta Kappan, 2005, 87(2), 104-111.
- [27] ISO FDIS 9241-210, Ergonomics of human system interaction - Part 210: Human-centered design for interactive systems, International Organization for Standardization (ISO), 2010.
- [28] S., Thiagarajan, Ask Thiagi. Thiagi Game Letter, 1998, 1(4), 6.
- [29] A. Tychsen, "Crafting user experience via game metrics analysis," Proc Workshop paper, NORDICHI 2008.
- [30] A. Waraich, "Using narrative as a motivating device to teach binary arithmetic and logic gates," Proc SIGCSE Conference on Innovation and Technology in Computer Science Education, ACM, 2004.
- [31] G. Yannakakis, and J. Hallam, "Towards Optimizing Entertainment in Computers Games," Applied Artificial Intelligence, vol 21, 2007, pp. 933-97, doi:10.1080/08839510701527580
- [32] A. Ibrahim, F.L. Gutiérrez, J.L. González Sánchez, and N. Padilla Zea, "Playability Design Pattern in Educational Video Game," Proc the 5th European Conference on Games Based Learning (ECGBL), October 2011.
- [33] Y. Li, "To Play Words Based on Video Identifying," Proc Second International Workshop on Education Technology and Computer Science, IEEE , pp.597-599, 2010.
- [34] F. Bellotti, R. Berta, A. De Gloria, and L. Primavera. "Enhancing the educational value of video games," Computer in Entertain, vol 7, 2, 2009, doi: 10.1145/1541895.1541903
- [35] A. Ibrahim, F.L. Gutiérrez, J.L. González Sánchez, and N. Padilla Zea, "Playability Design pattern in educational video game," In Student Usability in Educational Software and Games: Improving Experiences, C.S. González, Eds. Hershey: IGI Global. In press.
- [36] K. Ash, Balancing Fun and Learning in Educational Games, [http://blogs.edweek.org/edweek/DigitalEducation/2011/08/balancing\\_fun\\_and\\_learning\\_in.html](http://blogs.edweek.org/edweek/DigitalEducation/2011/08/balancing_fun_and_learning_in.html). (accessed on 26/11/11).
- [37] The Good and Bad Effects of Video Games, <http://www.raisesmartkid.com/3-to-6-years-old/4-articles/34-the-good-and-bad-effects-of-video-games> . (accessed on 26/11/11).
- [38] Mavis Beacon, <http://www.mavisbeacon.com>. (accessed on 26/11/11).
- [39] The Education Arcade, <http://www.educationarcade.org/node/357>. (accessed on 26/11/11).

# User Experience: Buzzword or New Paradigm?

Dominique L. Scapin, Bernard Senach, Brigitte Trousse, and Marc Pallot  
 INRIA Research Team AxIS  
 Rocquencourt and Sophia-Antipolis, France  
 {dominique.scapin, bernard.senach, brigitte.trousse, marc.pallot}@inria.fr

**Abstract - This paper explores User Experience, a rather novel and popular view on human-computer interaction, through an extensive review of the literature. After introducing its polysemous nature, this paper describes the origins of User Experience, its scope, components and various definitions. Then, User Experience methods are surveyed, distinguishing processes, frameworks, and specific methods. The conclusion identifies a set of issues about the needs for increased User Experience maturity.**

**Keywords - User Experience, Usage, HCI, New Paradigm, Hedonic, Pragmatic, Methods.**

## I. INTRODUCTION

From the early days of ergonomics and HCI (Human-Computer Interaction), user experience really meant user characteristics in terms of knowledge, skills, know-how, "savoir faire" through learning and practice. Usual distinctions were made between expert and novice, between domain expertise and computer expertise, etc.

In recent years, the meaning of User Experience, also named by its acronym UX, has changed, particularly under the influence of consumer products and marketing strategies. In 2004, Philips modified its advertisement from "Let's make things better" to "Sense and Simplicity". In 2005, Apple used "Enjoy uncertainty" to advertise its iPod. In 2006, Nike started a "Joint Product eXperience" campaign.

UX has become very popular in industry followed later on by several scientific communities. A growing number of networks, shared blogs, and wikis, have been initiated, mainly for computer industry professionals [1]. New educational opportunities are opening, as well as new UX-related jobs: conferences employment boards, professional email lists, etc., show a large progression in job offers that include UX requirements [2], [3].

However, UX has several meanings, with a varying and complex coverage of topics and issues. What is really UX, where does it come from, what does it mean, is it really that new? It is the purpose of this paper to uncover the complexity and underlying mechanisms of UX, particularly from a conceptual and methodological point of view.

This paper presents the UX origins, scope, components and various definitions. Then, UX methods are surveyed, distinguishing processes, frameworks, and specific methods. The conclusion identifies a set of issues about the needs towards increased UX maturity.

## II. UX ORIGINS

A good and humorous start is to look at Tom Stewart's (Chair of ISO TC 159/ SC4) company web site [4] (July 2009): "*The study of the relationship between people and technology has been called a variety of names over the years from computer ergonomics, human computer interaction and usability to, more recently, human-centred design and UX. The term user experience is now widely used, ... Personally, I do not really care what this area is called ... So I now use the term user experience to describe what I work on ...*"

Looking way back in terms of UX origin, an historical link can be established with [5]. This philosophical contribution about art contained already some of the UX pragmatic and holistic orientations of today.

In HCI, with the view that UX may just be another label, the origins of UX can be dated quite early, towards the start of Human Factors, during World War II, sometimes even earlier [6]. In the 80's-90's UX can be related to the rise of UCD (user-centered design) [7]. However the job of UX architect was then very limited in scope.

Indeed the question is: what is the difference between good old usability and UX, and what does it bring as new methods and results? This will be discussed in the definition Section IV, and in the methods Section V.

One of the first papers looking at UX with a very wide view can be found in [8]. At the time, based on several philosophical views, the author identified two types of experiences in user-product interactions: a satisfying experience, which is a process-driven act that is performed in a successful manner, and a rich experience, which has a sense of immersive continuity and interaction, and may be made up of a series of satisfying experiences.

Another line of thought regarding UX experience has been the "business view" [9], very much related to customer satisfaction and loyalty. A positive experience means a happy customer who returns again. Designers of software systems and web services have been digging deeply into how they might generate a positive UX. They are moving beyond anecdotes about excellent examples of UXs and are developing design principles

To sum up, the concept of UX is wide due to a holistic (preferably satisfying) experience and to the business point of view, i.e., the selling of products. Both points of view mean also that, unlike classical ergonomics, it concerns usability (which partly includes satisfaction, see ISO 9241-

Part 11) but a much stronger focus on non-work software, consumer products, and leisure applications.

III. UX BACKGROUND AND COMMUNITIES

What can at least be said about UX is that it corresponds definitely to the multi-disciplinary needs of industry [10]. Actually, a number of groups, communities and associations are listed in [11]. Some even talk about UX evangelism within organizations [12], where, at times, it is claimed that a well designed product should market itself, and that money is best spent on design and internal evangelism. However, collaboration between many different professional organizations might not be optimal [13].

Distinguishing circumstantial experience, long term experience, and co-experience, [14] identifies the issues and domains/ professions that concur to the various views of UX.

To sum up on the issue is what does UX covers, and to which communities does it belong, our view is that, as we will see later in the definitions and methods, the claim that it should be holistic and time dependant will require a much wider definition of UX (in terms of scientific backgrounds), as well as some return to the basics (to incorporate UX through time), and more efforts in defining new methods (and more importantly coordination between methods), as the coverage and novelty of UX specific methods seem currently quite limited.

To show roughly where UX work is, Table 1 provides a geographical assignment based on the authors and labs. found in our literature survey.

TABLE I.

Areas	Countries	Nb. Individual Contributors	Nb. Research Teams
<b>North America</b> - 60 individuals - from 30 labs	USA	58	29
	Canada	2	1
<b>Europe</b> - 82 individuals - from 53 labs	UK	23	14
	Netherlands	15	6
	Finland	14	6
	Germany	12	7
	Sweden	7	3
	Iceland	2	2
	Switzerland	2	2
	France	2	2
	Italy	2	7
Greece	1	3	
<b>Others</b> - 5 individuals - from 3 labs	Australia	3	1
	Israel	1	1
	Algeria	1	1

One can observe a large concentration of UX groups and labs in Northern Europe and the US, but very few from elsewhere, but they may have been overlooked.

IV. UX DEFINITIONS

To better characterize UX, it is useful to look at some of most cited definitions in the literature among the many currently available ones.

A very official one comes from ISO: ISO 9241-210 (2010) "*person's perceptions and responses resulting from*

*the use and/or anticipated use of a product, system or service*" with explanatory notes saying " *User experience includes all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours and accomplishments that occur before, during and after use.*" and " *User experience is a consequence of brand image, presentation, functionality, system performance, interactive behaviour and assistive capabilities of the interactive system, the user's internal and physical state resulting from prior experiences, attitudes, skills and personality, and the context of use.*" and " *Usability, when interpreted from the perspective of the users' personal goals, can include the kind of perceptual and emotional aspects typically associated with user experience. Usability criteria can be used to assess aspects of user experience.*".

Another definition comes from UPA [15], with acronym UE instead of UX: "*Every aspect of the user's interaction with a product, service, or company that make up the user's perceptions of the whole. User experience design as a discipline is concerned with all the elements that together make up that interface, including layout, visual design, text, brand, sound, and interaction. UE works to coordinate these elements to allow for the best possible interaction by users.*"

For both UX and UE definitions, major criticisms concern its impreciseness and the wide gap between practitioners and academics in their understanding [16]. A widely accepted, shared understanding of UX is still lacking. While UX seems ubiquitous in industry, a closer look reveals that it is treated mainly as a synonym of usability and user-centred-design. Academics, however, emphasize the differences between traditional usability and UX.

UX is viewed as a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g., complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g., organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.).

A survey [17] later showed that it was hard to gain a common agreement on the nature and scope of UX. However most respondents (275 researchers and practitioners from academia and industry) agree that UX is dynamic, context-dependent, and subjective, that UX is something individual (instead of social) that emerges from interacting with a product, system, service or an object. However, the issues of experiencing anticipated use and the object of UX are less consensual.

To sum up, while the concept of UX magnifies the issue of a subjective view, which does not restrict itself (unlike usability) to satisfaction of use, but to a much wider view on basic human needs, including aspects that are usually more related to marketing, art, communications, and organizational psychology.

While the focus is on issues such as perception, affects/ internal states, holistic, through time, it also does not just concern the product or service, but "interactions with the company", about the way it looks and is remembered [18].

In terms of scope, UX is obviously not restricted to work systems, as it concerns many subjective aspects beyond

performance. However, UX can still apply in work systems, with the view that it can motivate people and improve work practices [19], or support extrinsically motivated experiences complementary to intrinsically motivated experiences [20].

## V. UX METHODS

After looking at scope and definitions, a number of questions arise: how UX is considered in the software process (including business oriented), what general UX design and evaluation frameworks are proposed, and what are some of the UX methods.

### A. Software Process and Business

The software process, its lifecycle is usually an item that gets attention once the domain is mature enough, which might not be the case for UX yet. However a few heterogeneous contributions have arisen, from detailed methodology to simple case studies.

For UX in the process, the challenge is there, as mentioned in [21]. The question is how industry and manufacturers manage to successfully get a UX idea into and through the development cycle? That is, to develop and sell it in the market within the right timeframe and with the right content. Proposals are made through UTUM (UIQ Technology Usability Metrics).

The issue is whether or not a UX perspective changes the process: obviously yes if goals are to incorporate the issue of selling the right thing at the right time. This is illustrated in a study by [22] that describes the kind of UX measurements that are said to be useful in different parts of one particular organization. This email survey led to some distinctions between product pre-purchase, first use, and upgrade.

Of course, one should also consider the variations due to the purpose of the product/service: knowledge, buying off the shelf, comparing, design, redesign, establishing conformance, etc.

In terms of development process, using scrum, an agile programming methodology, a case study [23] led to a few recommendations about the workflow, roles and responsibilities of UX and cross-functional team in an hybrid agile environment.

Concerning the links between UX and business, a case study [24] reported some experience (website, standards, training) in creating a UX culture, focusing on distributed design teams, and interactions with vendors and business processing outsourcing efforts.

Other views of that nature can be found in [25] about organizational views specific to one company and in [26] about shared intelligence.

While mainly on usability, another study [27] offers a view joining task impact and business impact. A similar distinction is made in [28] between Strategic usability measures (business) and Operational usability measures (user performance).

### B. UX design and evaluation frameworks

The intent to separate design and evaluation studies seems unpractical as most studies do not make such

distinctions, even though it seems there are more design aspects in UX studies than in usability studies.

In [29], a theoretical design framework is introduced as "Product Ecology" and offers means of selecting relevant research methods. The methods are not really new (observation, log-files, diaries, group interactions, as well as participatory design, cultural probes, etc.) but include explicitly social aspects.

In [30], a similar view stresses that experience goes beyond the artifact and actual use; it is a momentum and has a timeframe. It also points out that it is not possible to design an experience, but rather to design for experience. On the practical side, they report results on a television portal survey exploring motivation and expectations, findings and willingness to tell others, as well as emotions and attractiveness.

In [31], the focus is on ways to encourage UX designer participation. It describes CodePlex, a community website that hosts open source projects. The main four design concepts on which the tool is built are: foster ways to build trust, provide opportunities for merit, tools to support crossover of work activities, and UX workspace commensurate with best practices.

Other contributions, such as [32] propose basically to extend most usability methods selection criteria to UX.

As for most new endeavors, new domains of investigation, many other proposals remain theoretical, or ad hoc, and rarely provide assessments from users, or clients. Also, they usually are in the form of opinions, rather than empirical evidence (in use, or re-use, or simply purchase).

In addition, it is worth mentioning that some biases have been already identified. For instance, in [33], two experimental studies support the idea of a framing effect, showing that the same information can lead to different judgments and decisions according to whether it is presented in a positive or in a negative frame. In addition, as external sources of judgment may infer (reputation, recommendations), it seems useful to investigate further social networks in their usage.

### C. Specific UX methods

This section gathers the contributions that deal explicitly with individual UX methods (i.e., not methodologies), with a few UX dimensions or methods comparisons. Also two areas that have inspired UX work are mentioned: software for games and children.

#### 1) Individual UX dimensions and methods:

Concerning **aesthetics** measures, reference work can be found in [34], and [35]. Most measures are based on classical psychological or social methods, but include also physiological measurements, such as heart rate, galvanic skin response, pupillometry and eye tracking.

Concerning **emotions**, work on measuring emotions can be found in [36], [37], [38]. However, the link between theory and practice is yet weak, as pointed out in [39], and it has the additional constraint to include the multidimensional nature of emotions (behavior, feelings), and its continuous nature, which implies lengthy and multiple measures.

Concerning **UX recording/ observing**, a system was built for the UX team at a major Internet company [40]. It simply states that the ability to watch high-definition study videos live from anywhere on the network led to a dramatic increase in the number of observers who directly experience their end users, but there are useful hints for recording and observation.

Concerning **Questionnaires**, the literature is richer. For instance, a questionnaire in [41] contains nine items to measure the occurrence of three core human needs taken from Self-Determination Theory: autonomy, competence and relatedness. Many others do exist (many ad hoc, internal, not tested) that stimulated the need for selecting an appropriate scale for UX in [42]. The main difficulty seems to be the lack of explicit and transparent descriptions of psychometric scale development. Also, specific to the domain of emotions [43], a list of 10 emotion words, five positive and five negative was extracted from a cluster analysis of resulting data obtained in a research prompting users in an evaluation of 6 websites.

Concerning **Heuristics**, an attempt was made in [44] to built and discuss how well heuristics covered the positive and negative service UX evaluation findings, and how the heuristics and the expert evaluation approach of UX should be developed further. As for usability heuristics, standard principles, or ergonomics criteria, it may need a while to be fully developed and assessed, together with the increased knowledge of UX and associated recommendations based on sound scientific grounds.

### 2) Comparison of UX dimensions and methods:

A comparison of pragmatic vs. hedonic in [45] showed that in a promotion focus (concern for safety and the avoidance of negative outcomes) participants rated an hedonic mp3-player as more appealing and chose it more frequently compared to individuals in a prevention focus (concern for personal growth and the attainment of positive outcomes). Reverse results, albeit not as strong, were found for the evaluation and choice of a pragmatic mp3-player. This supports the idea that the perceived quality of interactive products can be roughly divided into instrumental, task-related, pragmatic attributes (e.g., usefulness, usability) and non-instrumental, self-referential, hedonic attributes (e.g., novelty, beauty). Along the same lines, a study [46] showed again relationships between type of tasks and types of measures, i.e., more or less hedonic value depending on the task, confirming that pragmatic issues get high scores for a task-oriented software.

### 3) Inspiring areas: UX in games and for children:

Regarding digital games, a study [47] explored the variety of experiences (i.e., positive and negative) that are received from playing. The results suggested that UXs are versatile in nature but they consist of four major constructs: cognition, motivation, emotion and focused attention. It also pointed out the role of gender. From the methodological side, it suggests that current technologies are not advanced enough to reveal the vast and rich amount of details in experiences. This is a reason for the unavoidable need to still approach

individuals' experiences by using traditional methods such as interviews and questionnaires. Another study [48] reviews the different elements of the gaming experience and their relation to other concepts within HCI. One advice, focusing on ownership, is that normal applications should learn from games, i.e., to pay attention to what the user is trying to do, and help the user make the tool his own.

Regarding **children**, a paper [49] showed how studying children's drawings can be an evaluation tool for capturing their experiences of different novel interfaces. Usability and UX factors: Fun (F), Goal Fit (GF) and Tangible Magic (TM) were included in the coding scheme. An interesting outcome is a correlation between usability and UX. Another paper [50] in an experiment using a Smileyometer from a Fun Toolkit, supported the idea of a difference between expectations and actual experience with children.

## VI. CONCLUSION

In the market place, UX has become a major component particularly for new computer devices, mobiles and internet.

The literature on UX is quite variable in its nature. Not all contributions gathered in our literature survey have been retained, the reasons being: vague statements, redundancies, magazine-like articles, even sometimes advertisement-like articles, and very few studies with empirical findings.

One area in which there is a lot of debate is who owns the domain and how can it be promoted. This is a usual characteristic of multidisciplinary domains in the making. This is interesting as it offers lots of opportunities for collaborative research to shape up the future.

An obvious need is some converging, non-polysemous, agreed upon definitions of UX that cover the various domains and territories that are involved. An effort to distinguish the concepts of quality in use, actual usability and user experience is proposed in [51].

An even more important need is the improvement of UX processes, methods, and tools. Of course, a large part of existing usability, accessibility, and marketing methods can be applied, or rather should be applied with the view that UX encompasses all aspects of user interactions with products and services. Several contributions, in terms of research and practice, can be found in [52].

Some methods (e.g., questionnaires, interviews, etc.) need just extension to more subjective areas of emotions, branding, etc. However, the nature of UX being very subjective, context-, and time-dependant, not all current processes, methods, and tools apply well, and many areas of UX involvement are not covered, whether positive or negative.

Even though, UX is still sometimes just viewed as an extension of usability, its future may really correspond to a paradigm evolution rather than simply a buzz word. The evolution is not drastic, but it adds complexity by considering more user areas than traditional usability.

In addition, UX aims at all venues of everyday life, at products and services, not only at work situations. In all cases UX attempts to include both pragmatic and hedonic goals, viewing the user from many angles: political, social,

marketing, art, as well as physiology, psychology, anthropology, etc.

Indeed, UX lead to new keywords: from "ergonomics" to "marketing" and "art", from "safety, efficiency" to "pleasure, enjoyment, engagement", from "design" to "innovation", from "lab & field studies" to "living labs", from "user centered" to "co-design".

The important issue is that for data gathering, it is changing as well: from speed, task goals achievement ratio, number of errors, etc., to perceived quality, goodness, engagement, seriousness/fun, etc. It is very important for well grounded UX future that the nature of data categories to gather is well defined, structured, and coordinated, together with improved methods.

The last question is to better understand the relationship between UX and usage. This question, even though not specific to UX, is particularly important as UX involves time. This has to do with predictive vs. actual use. Our view is that a distinction should be made within UX between:

- UX as a concept covering widely all the aspects described above (including cognitive perception and representation by the users).
- UX as a result or state, predicted from theory or knowledge-based methods, or even actually measured at a specific time, in a particular context (for instance in an experiment).
- UX as usage, i.e., the actual use as it can be monitored, surveyed, assessed. This corresponds to actual operations, actions and perceptions of the users through time (directly observable or not).

This question will be debated within various domains and communities, including the emerging experiential research and innovation of Living Labs aimed at leading user communities towards group cognition and collective intelligence based on accumulated experience knowledge that enriches technology platforms. Indeed, today, users/citizens are rather considered as potential co-creators and experimenters that generate new ideas, play with them, feel, sense and interact within real scenarios and prototyped products/services [53], [54].

To end as started, on another humorous note: "Usability wants us to die rich. UX wants us to die happy" [55].

#### REFERENCES

[1] B. Helfer, B. "ACM SIGGRAPH user experience initiatives". In CHI '05 Extended Abstracts on Human Factors in Computing Systems (Portland, OR, USA, April 02 - 07, 2005). CHI '05. ACM, New York, NY, 1085-1086.

[2] [http://twitter.com/ia\\_uxjobs](http://twitter.com/ia_uxjobs) (Retrieved September 2011)

[3] <http://bx.businessweek.com/user-experience-ux/jobs/> (Retrieved September 2011)

[4] <http://www.system-concepts.com/articles/usability-articles/2008/usability-or-user-experience-whats-the-difference.html> (Retrieved september 2011)

[5] J. Dewey. "Art as experience". New York: Capricorn Books, 1934.

[6] W. Jastrzębowski "Rys ergonomji czyli nauki o pracy, opartej na prawdach poczerpniętych z Nauki Przyrody" (The Outline of Ergonomics, i.e. Science of Work, Based on the Truths Taken from the Natural Science) (1857)

[7] D. Norman, J. Miller, A. Henderson. "What You See, Some of What's in the Future, And How We Go About Doing It: HI at Apple Computer". Proc. of CHI 1995, Denver, Colorado, USA.

[8] J. L. Forlizzi. "Designing for Experience: An Approach to Human-centered Design". Master of Design in Interaction Design, Department of Design, College of Fine Arts, Carnegie Mellon University, May 1997.

[9] P. Tobias and D. S. Spiegel. "Is Design the Preeminent Protagonist in User Experience?" Ubiquity 2009, May (May. 2009).

[10] K. Instone, "User experience: an umbrella topic" In CHI '05 Extended Abstracts on Human Factors in Computing Systems (Portland, OR, USA, April 02 - 07, 2005). CHI '05. ACM, New York, NY, 1087-1088.

[11] P. Sherman and W. Quesenbery, "Engineering the user experience: UX and the Usability Professionals' Association". Interactions 12, 3 (May. 2005), 38-40.

[12] L. Kowalski, C. Thompson., T. Chi, D. Mc Cormick, O. Vasnaik, and P. Heller. "What would you do with a 1 million dollar user experience marketing budget?: internal vs. external user experience evangelism." In CHI '08 Extended Abstracts on Human Factors in Computing Systems (Florence, Italy, April 05 - 10, 2008). CHI '08. ACM, New York, NY, 2249-2252.

[13] R. I. Anderson. "Meeting the needs of the "user experience" professional". In CHI '05 Extended Abstracts on Human Factors in Computing Systems (Portland, OR, USA, April 02 - 07, 2005). CHI '05. ACM, New York, NY, 1158-1159.

[14] J. Forlizzi and K. Battarbee. Understanding experience in interactive systems. In Proc. of the 5th Conference on Designing interactive Systems: Processes, Practices, Methods, and Techniques (Cambridge, MA, USA, August 01 - 04, 2004). DIS '04. ACM, New York, NY, 261-268.

[15] <http://www.usabilitybok.org/glossary> (Retrieved September 2011)

[16] M. Hassenzahl. "User experience (UX): towards an experiential perspective on product quality". In Proc. of the 20th international Conference of the Association Francophone D'interaction Homme-Machine (Metz, France, September 02 - 05, 2008). IHM '08, vol. 339. ACM, New York, NY, 11-15.

[17] E. C. Law, V. Roto, M. Hassenzahl, A. P. Vermeeren, and J. Kort, J. " Understanding, scoping and defining user experience: a survey approach". In Proc. of the 27th international Conference on Human Factors in Computing Systems (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 719-728.

[18] F. Sampson. Brand UX. interactions 12, 4 (Jul. 2005), 10-11.

[19] L. Norros, and P. Savioja. "Use Experience in Systems Usability Approach" In Proc. of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM). Reykjavik, Island, (2008) pp. 45-48.

[20] A. Lesage and T. Dorta, " Au-delà de l'utilisabilité: l'autotélie". In Proc. of the 20th international Conference of the Association Francophone D'interaction Homme-Machine (Metz, France, September 02 - 05, 2008). IHM '08, vol. 339. ACM, New York, NY, 147-150.

[21] M. Hellman, and K. Rönkkö, "Is User Experience Supported Effectively in Existing Software Development Processes?" In Proc. of COST294-MAUSE Workshop on Valid Useful User Experience Measurement (VUUM). Reykjavik, Island, (2008) pp. 32-37.

[22] P. Ketola, and V. Roto, "Exploring User Experience Measurement Need"s. In Proc. of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM). Reykjavik, Island, (2008) pp. 23-26.

[23] M. Budwig, S. Jeong, and K. Kelkar. "When user experience met agile: a case stud"y. In Proc. of the 27th international Conference on Human Factors in Computing Systems (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 3075-3084.

[24] B. Leadley, H. Pao, and S. Douglas. "Creating a user experience culture at a non-software company". In Proc. of the 2005 Conference on Designing For User Experience (San Francisco, California,



- November 03 - 05, 2005). Designing For User Experiences, vol. 135. AIGA: American Institute of Graphic Arts, New York, NY, 38.
- [25] D. Rosenberg. "Introducing the 360° View of UX Management". *Interactions*, interactions 14, 3 (May. 2007),, pp. 23-24.
- [26] J. Innes. "Defining the user experience function: innovation through organizational design". *interactions* 14, 3 (May. 2007), 36-37.
- [27] L. Gorlenko and P. Englefield. "Usability error classification: qualitative data analysis for UX practitioners". In *CHI '06 Extended Abstracts* (Montréal, Québec, Canada, April 22 - 27, 2006). CHI '06. ACM, New York, NY, 803-808.
- [28] T. Jokela. "A Two-Level Approach for Determining Measurable Usability Targets". In *Proc. of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM)*. Reykjavik, Island, (2008) pp. 56-59.
- [29] J. Forlizzi. The product ecology: Understanding social product use and supporting design culture. *International Journal of Design*, 2(1), 11-20. (2007).
- [30] E. R. Oppelaar, E. Hennipman, and G. van der Veer. "Experience design for dummies". In *Proc. of the 15th European Conference on Cognitive Ergonomics: the Ergonomics of Cool interaction* (Funchal, Portugal, September 16 - 19, 2008). J. Abascal, I. Fajardo, and I. Oakley, Eds. ECCE '08, vol. 369. ACM, New York, NY, 1-8.
- [31] P. M. Bach, R. DeLine, and J. M. Carroll. "Designers wanted: participation and the user experience in open source software development". In *Proc. of the 27th international Conference on Human Factors in Computing Systems* (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 985-994.
- [32] N. Bevan. "Classifying and Selecting UX and Usability Measures". In *Proc. COST294-MAUSE Workshop: Valid Useful User Experience Measurement (VUUM)*. Reykjavik, Island, (2008) pp. 13-18.
- [33] J. Hartmann, A. De Angeli, and A. Sutcliffe. Framing the user experience: information biases on website quality judgement. In *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy, April 05 - 10, 2008). CHI '08. ACM, New York, NY, 855-864.
- [34] M. Hassenzahl, G. Lindgaard, and N. Tractinsky. "The study of visual aesthetics in human computer interaction", Proc. Dagstuhl Seminar 08292, 10-13 July (2008), Dagstuhl, Germany.
- [35] J. Hartmann, A. Sutcliffe, and A. Angeli. "Towards a theory of user judgment of aesthetics and user interface quality." *ACM Transactions on Computer-Human Interaction (TOCHI)* 15, no. 4(2008) : 1-30.
- [36] R. Hazlett and J. Benedeck. "Measuring emotional valence to understand the user's experience of software". *International Journal of Human-Computer Studies*, Vol. 65, 2007, pp. 306-314.
- [37] K. Isbister, K. Höök, J. Laakso, and M. Sharp. "The sensual evaluation instrument: Developing transcultural self-report measure of affect". *International Journal of Human-Computer Studies*, Vol. 65, 2007, pp. 315-328.
- [38] M. Isomursu, S. Tähti, S. Väinämö, and K. Kuutti. "Experimental evaluation of five methods for collecting emotions in field settings with mobile applications". *International Journal of Human-Computer Studies*, Vol. 65, 2007, pp. 404-418.
- [39] N. Gauducheu. "Mesurer les émotions de l'utilisateur: quels fondements pour une démarche d'évaluation des systèmes interactifs?". In *Proc. of the 21st international Conference on Association Francophone D'interaction Homme-Machine* (Grenoble, France, October 13 - 16, 2009). IHM '09. ACM, New York, 183-192.
- [40] M. LaRosa, D. Poole, and R. Schusteritsch. "Designing and deploying usetube, google's global user experience observation and recording system". In *Proc. of the 27th international Conference on Human Factors in Computing Systems* (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 2971-2986.
- [41] M. Hassenzahl, S. Diefenbach, and A. Saran, "Needs, affects, and interactive products – Facets of user experience.". *Interacting with Computers*, 22 (5) pp. 353-362.
- [42] W. Green, G. Dunn, and J. Hoonhout, J. "Developing the Scale Adoption Framework for Evaluation". In *Proc. of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM)*. Reykjavik, Island, (2008) pp. 49-55.
- [43] H. Petrie and C. Harrison. Measuring users' emotional reactions to websites. In *Proc. of the 27th international Conference on Human Factors in Computing Systems* (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 3847-3852.
- [44] Väänänen-Vainio-Mattila, K. and M. Wäljas, "Developing an expert evaluation method for user eXperience of cross-platform web services". In *Proc. of the 13th international MindTrek Conference: Everyday Life in the Ubiquitous Era* (Tampere, Finland, Sept. 30 – Oct. 02, 2009). MindTrek '09. ACM, New York, NY, 162-169.
- [45] M. Hassenzahl, M. Schöbel, and T. Trautmann, "How motivational orientation influences the evaluation and choice of hedonic and pragmatic interactive products: The role of regulatory focus". *Interact. Comput.* 20, 4-5 (Sep. 2008), 473-479.
- [46] J. Isleifsdottir and M. Larusdottir. "Measuring the User Experience of a Task Oriented Software". In *Proc. of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM)*. Reykjavik, Island, (2008) pp. 97-102.
- [47] . Komulainen, J. Takatalo, M. Lehtonen, and G. Nyman, G. "Psychologically structured approach to user experience in games". In *Proc. of the 5th Nordic Conference on Human-Computer interaction: Building Bridges* (Lund, Sweden, October 20 - 22, 2008). NordiCHI '08, vol. 358. ACM, New York, NY, 487-490.
- [48] E. H. Calvillo Gómez, P. Cairns, and A. L. Cox. "From the gaming experience to the wider user experience". In *Proc. of the 2009 British Computer Society Conference on Human-Computer interaction* (Cambridge, United Kingdom, September 01 - 05, 2009). British Computer Society Conference on Human-Computer Interaction. British Computer Society, Swinton, UK, 520-523.
- [49] D. Y. Xu, J. C. Read, G. Sim, B. McManus, and P. Qualter, P. "Children and 'smart' technologies: can children's experiences be interpreted and coded?". In *Proc. of the 2009 British Computer Society Conference on Human-Computer interaction* (Cambridge, United Kingdom, September 01 - 05, 2009). British Computer Society Conference on Human-Computer Interaction. British Computer Society, Swinton, UK, 224-231.
- [50] J. Read. "Is what you see what you get? Children, Technology and the Fun Toolkit". In *Proc. of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM)*. Reykjavik, Island, (2008) pp. 67-71.
- [51] P. Lew, L. Olsina, and L. Zhang, "Quality, Quality in Use, Actual Usability and User Experience as Key Drivers for Web Application Evaluation", in *Proc. ICWE*, 2010, pp.218-232.
- [52] E. Law and P. Van Schaik. (Eds.) Special Issue Interacting with Computers, (2010) Vol. 22, Issue 5, pp. 313-438.
- [53] <http://www.ami-communities.eu/wiki/>, <http://www.openlivinglabs.eu/> (Retrieved September 2011)
- [54] M. Pallot, B. Trousse, B. Senach, and D. L. Scapin. "Living Lab research landscape: From user centred design and user experience towards user cocreation". In *1st. European Living Labs Summer School, Collaborative Innovation through Living Labs*, Cité des Sciences, Paris, France, august 2010
- [55] M. Hassenzahl and V. Roto, V. "Being and Doing - A perspective on user experience and its measurement". *Interfaces Magazine* 72, British HCI Group. (2007)

# Enhancing Automatic Detection of Frustration Induced During HCI with Moment-based Biosignal Features

Dimitris Giakoumis<sup>1,2</sup>, Dimitrios Tzovaras<sup>1</sup>

<sup>1</sup>Informatics and Telematics Institute  
Centre for Research and Technology Hellas  
Thessaloniki, Greece  
dgiakoum@iti.gr, tzovaras@iti.gr

George Hassapis<sup>2</sup>

<sup>2</sup>Dept. of Electrical and Computer Engineering  
Aristotle University of Thessaloniki  
Greece  
ghass@auth.gr

**Abstract**—Enhancing HCI systems with the capability to detect user’s frustration and respond appropriately is a significant challenge. In this line, biosignal features based on the theory of orthogonal Krawtchouk and Legendre moments are assessed in the present work over their ability to enhance accuracy in automatic detection of frustration, which is induced through HCI, during video-game playing. Experimental evaluation, conducted over a multi-subject dataset over frustration detection showed that conventional features, typically extracted from Galvanic Skin Response and Electrocardiogram in the past, achieved correct classification rate (CCR) of 83.59%. Fusing these conventional features with moment-based ones extracted from the same modalities resulted to significantly higher accuracy, at the level of 93%. Furthermore, moment-based features lead also to over 10% increase in CCR when the aim was to identify both bored and frustrated cases, within a 3-class affect detection problem.

**Keywords**- automatic frustration detection, biosignals, moment-based features, video game-playing

## I. INTRODUCTION

Negative emotional states like frustration are likely to be induced during Human - Computer Interaction (HCI). Frustration is an emotional state commonly associated with anger. During HCI, it can cause a negative disposition of the user towards the machine [1]. In the context of video games, frustration is typically induced when game difficulty is in mismatch with the capabilities and/or preferences of the player [2]. It can lead to player’s disassociation from the game, dissatisfaction and resign. This is a case that may occur even in modern video-games, which, although carefully designed, do not take into account the player’s current emotional state and its specifics [2].

Future game-playing systems can be augmented with the capability to automatically detect the player’s affective state and monitor user experience [3]. When needed, these systems will be capable to adjust playing context appropriately [4], so as to maintain entertainment through a closed biocybernetic loop [5]. Frustration is an emotional state that can play a key role in this context [6], since as it has been shown in the past, machines responding to player’s frustration can lead to improved gaming experience [1,2]. However, a pre-requisite for building such future game-

playing systems, is to provide machines the capability to detect frustration effectively.

### A. Related Work

During the past years, significant progress has been made in the field of automatic affect detection (e.g. [9,10,21]). This progress is important for a large variety of future HCI applications, ranging e.g. from affective games [7] to affective intelligent tutoring systems [12] that can be based on emotion sensitive e-learning models [8]. In this context, important efforts have been made so far towards enabling machines to automatically detect frustration. These studies utilized biosignals [11,12,14], video [12], or other data types [11,12] recorded during frustration induction, so as to build classifiers appropriate for detecting this negative emotional state.

Features extracted from the Galvanic Skin Response (GSR) and Blood Volume Pulse (BVP) were used in [14], leading to frustration detection accuracy of 67.39% among a multi-subject (MS) dataset, having a computer game as stimuli. In [12], the focus was towards an affective learning companion, and frustration was predicted with accuracy of 79% (MS), by utilizing features extracted from a face tracker, a posture sensing chair, a pressure mouse and GSR. Physiological features merged with contextual ones, extracted during students’ interaction with a tutoring system, lead in [11] to 88.8% frustration detection accuracy. In [15], using features extracted from biosignal (GSR, temperature and heart rate) modalities, frustration was recognized from five further emotion classes with accuracy of 78.3%. Focusing on HCI in respect of video games, biosignal (e.g. GSR and Respiration) features were found to correlate in [3] with frustration induced from video-game playing, whereas in [4] frustration was detected from biosignals and gameplay data with accuracy around 85%.

From the above it is clear that in general, biosignals have good potential towards automatic frustration detection. However, although affect detection has significantly advanced during the recent years [16], frustration detection accuracy levels as well as in general emotion recognition (ER) ones have remained relatively limited, i.e. only rarely exceeding 90%. Therefore, evident is the need for new biosignal processing techniques, which will lead to more effective ER systems. Working towards this direction, [13] proposed biosignal features extracted from GSR and Inter-

Beat-Intervals (IBI) time series (calculated from the Electrocardiogram (ECG)), which are based on the theory of orthogonal Legendre [17] and Krawtchouk [18] moments. These moments have been widely used in the past for the purposes of image analysis and reconstruction. Such a typical example can be found in [18], where it was shown that based on weighted Krawtchouk moments, effective image reconstruction and object recognition can be achieved. However, before [13], features based on Legendre and Krawtchouk moments had not been considered as an option in biosignals-based affect detection. The use of these features, together with conventional, not-moment based ones typically utilized in the past, was found in that work to significantly increase effectiveness of automatic boredom recognition.

**B. Contribution**

Boredom and frustration are both negative emotions of high significance in the context of HCI applications. Therefore, the present work aims to enhance effectiveness of automatic frustration detection, by using moment-based biosignal features. Moreover, taking a step further from [13], we assess whether the moment-based features can deal with the problem of detecting whether the subject is getting both bored and frustrated during HCI. As explained in the following, it was found that combining moment-based features with conventional ones can significantly enhance the effectiveness of automatic frustration and also joint boredom/frustration detection, compared to the case where conventional features are used alone.

**II. BIOSIGNAL FEATURES EXTRACTION**

Various biosignal features were examined in the present study over their effectiveness in the given context. All features presented in the following were extracted from GSR and IBI time series recorded during rest periods and game-playing trials of the dataset described in Section 3.

**A. Conventional features**

A set of “conventional” features was first extracted from all game-playing trials. These features, summarized in Table I, have proved in the past capable to form the basis for systems targeting automatic frustration detection, and ER in general. More details regarding the specifics (e.g. formulas) for the extraction of these features in the present study can be found in [13].

TABLE I. FEATURES EXTRACTED FROM THE GSR SIGNAL AND THE INTER BEAT INTERVALS (IBI) TIME SERIES

Signal	Conventional Features Extracted
GSR	Mean, Standard Deviation (SD), 1 <sup>st</sup> derivative average, 1 <sup>st</sup> derivative RMS, Number of SCRs, Average Amplitude of SCRs, Average Duration of SCRs, Maximum Amplitude of SCRs, $\delta(\text{gsr})$ , $\delta_{\text{norm}}(\text{gsr})$ , $\gamma_{\text{norm}}(\text{gsr})$ , $f_d(\text{gsr})$
IBI	Mean, SD, LF/HF, RMSSD, pNN50, $\delta(\text{ibi})$ , $\delta_{\text{norm}}(\text{ibi})$ , $\gamma_{\text{norm}}(\text{ibi})$ , $f_d(\text{ibi})$

Moreover, all features of Table I were also extracted from only the first and last 10 seconds of each trial or resting

period; then, the ratio between each feature’s value calculated from the first 10 seconds to the corresponding value calculated of the last 10 seconds was extracted as an extra feature (marked in the rest of the paper with the extension “*FL*”). These ratios were calculated for all features that were applicable, similarly to [13]. In total, 37 conventional features were extracted, 9 from IBI, 12 from GSR, and 16 as the feature value ratio between the first and last 10 secs of each trial.

**B. Biosignal Features Based on the Theory of Moments**

Legendre moments [17] are based on projecting a signal onto Legendre polynomials, which form a complete orthogonal basis set defined over the interval [-1,1]. For a 1D discrete signal  $f(x_i)$ ,  $1 \leq i \leq N$ , the 1D Legendre moment of order  $p$  is given by:

$$L_p = \frac{2p+1}{N-1} \sum_{i=1}^N P_p(x_i) f(x_i) \tag{1}$$

where  $x_i=(2i-N-1)/(N-1)$  and  $P_p(x)$  is the  $p^{\text{th}}$  order Legendre polynomial given by:

$$P_p(x) = \frac{1}{2^p} \sum_{k=0}^{p/2} (-1)^k \frac{(2p-2k)!}{k!(p-k)!(p-2k)!} x^{p-2k} \tag{2}$$

where  $x$  belongs in the span [-1,1]. Legendre polynomials were calculated with appropriate recursive relation [13]. Legendre moments of orders 0-39 were calculated for the GSR and IBI signals (features *gsr\_LgXX* and *ibi\_LgXX* respectively, where *XX* is the moment order), taken from the first 25 seconds of each trial so as to ensure uniformity in the extraction process. Prior to feature extraction, signals were sub-sampled at 4Hz and normalized to their subject-specific global min and max values by  $\bar{X}(i) = (X(i) - X_{\min}) / (X_{\max} - X_{\min})$ , where  $X$  is either the GSR or IBI signal,  $X(i)$  is a GSR or IBI sample,  $X_{\min}$  and  $X_{\max}$  are the GSR or IBI signal’s min and max values recorded during all the specific subject’s game-playing trials. Only the first 40 orders were extracted as features; the use of higher ones would increase complexity and was not expected to provide added value. As shown in [10], these orders were capable to capture information conveyed through signal frequencies approximately up to 0.5Hz.

Krawtchouk moments are based on a set of orthonormal polynomials; the  $n$ -order Krawtchouk classical polynomials are defined as:

$$K_n(x; p, N) = \sum_{k=0}^N a_{k,n,p} x^k = {}_2F_1(-n, -x; -N; \frac{1}{p}) \tag{3}$$

where  $x, n=0,1...N$ ,  $N>0$ ,  $p$  belongs in the span (0,1) and  ${}_2F_1$  is the hypergeometric function [18]. Weighted Krawtchouk polynomials ( $\bar{K}_n(x; p, N)$ ) were introduced in [18]. For a 1D signal  $f(x_i)$  of length  $N$ , the weighted Krawtchouk moments  $\bar{Q}_n$  are defined as:

$$\bar{Q}_n = \sum_{i=1}^N \bar{K}_n(i-1; p, N-1) f(x_i) \tag{4}$$

where  $x_i=i-1$ . In our case  $p$  was taken equal to 0.5, in order for the region-of-interest of the feature extraction process to be centered at the half of each trial's first  $N$  samples. The 40 first Krawtchouk moments (0-39) were calculated with (4) for the GSR and IBI time series (features  $gsr\_KrXX$  and  $ibi\_KrXX$  respectively,  $XX$  is the moment order), by following the same specifics as in the afore-described Legendre moments case. The analysis was restricted to the first 40 orders; in this case information conveyed through signal frequencies approximately up to 0.8Hz was captured.

The moment-based feature variations proposed in [13] were also extracted and assessed over their effectiveness in the present work's context. These moment-based feature variations have the rationale of suppressing the static parameter of the original moments calculation; i.e. the area between the projection polynomial and the  $x$  axis, which is always identical. By using (5) and (6) instead of (1) and (4) respectively, these features are defined as:

$$L_p^{mod} = (2p + 1) \sum_{i=1}^N P_p(x_i)(f(x_i) - 1) \quad (5)$$

$$\bar{Q}_n^{mod} = \sum_{i=1}^N \bar{K}_n(i - 1; p, N - 1)(f(x_i) - 1) \quad (6)$$

Based on the first 40 Legendre polynomials, 40 features were extracted from GSR and IBI signals (features  $gsr\_Lg_{mod}XX$  and  $ibi\_Lg_{mod}XX$ ), by following the same procedure as in the original Legendre moment-based features case, and using (5) instead of (1). Similarly, by using (6) instead of (4), 40 further Krawtchouk-based features were extracted from each signal (features  $gsr\_Kr_{mod}XX$ ,  $ibi\_Kr_{mod}XX$ ).

### III. FRUSTRATION INDUCTION THROUGH REPETITIVE VIDEO GAME PLAYING

All aforedescribed features were extracted from biosignals recorded through the experimental process described in [13]. The specific experiment had the purpose of naturally inducing negative emotions like boredom to subjects, by the repetitive playing of the same video-game. The game utilized was an easy "3D Labyrinth" one. In each repetition (trial), the subject started from the same point and had to find the exit of the labyrinth, which was always located at the same place. The "3D Labyrinth" resembled on its gameplay basis to modern commercial games played by vast amounts of gamers worldwide (i.e. 3D-based first person role playing games). At the same time, the overall repetitive playing procedure lacked in all three of Malone's intrinsic qualitative factors for engaging game play (challenge, curiosity and fantasy) [19]. As a result, although at the beginning the game could be considered somewhat exciting, as soon as the subject had learned the shortest path to the labyrinth exit, boredom and negative emotions due to loss of interest were naturally induced.

Taking into account the appraisal theory [20], the main factor manipulated during the experimental session was novelty, the absence of which is a key factor for boredom induction. Furthermore, low novelty may result to the

induction of further emotions, such as irritation / cold anger. Therefore, it was rational to expect the appearance of frustration in subjects during the session, an emotional state that was monitored by self-reports (mid-trials questionnaires) throughout the experiment. After each trial, the subject answered a few questions directly assessing her/his emotional state. Among these questions were Likert-scaled (1-5, with labels in the range "Not at all"- "Very Much") ones regarding the self-assessment of boredom and frustration that the subject experienced during the last trial, as well as one asking whether s/he wanted to play the game again.

Data was collected from 19 subjects (14 male, 5 female) who frequently used computers in their work. These were between 23 and 44 years old, and their average age was 29. In total, 221 trials were recorded. The collected biosignals data was annotated as "Not Frustrated" (NF) or "Frustrated" (F) on the basis of the subjects answers to the frustration self assessment question. Each trial after which the answer to this question was "1" or "2" was labeled as belonging to the NF class. If this answer was "4" or "5", the trial was assigned to the F class. Trials after which the respective answer was "3" were excluded from further analysis. As a result, an annotated dataset (A) consisting of 195 trials, 149 belonging to the NF and 46 to the F class, was obtained. Moreover, one further annotated dataset (B) was deployed, formulating a 3-class ER problem, where trials were labeled as "not bored" (NB), "bored and not frustrated" (B/NF), or "bored and frustrated" (B/F). The idea behind dataset B was to evaluate the given biosignal GSR and IBI features over their capability to differentiate between cases of subjects who are 1) not bored, 2) bored, but not frustrated, 3) bored, to the extent where frustration has also appeared during HCI. For this purpose, all trials after which the subject's answer to the boredom self-assessment question was "1" or "2" (denoting absence of boredom) were annotated as NB. The rest of trials were annotated as B/NF or B/F, in respect to the answer to the frustration self-assessment question, similarly to the annotation of dataset A. Trials for which the answer either to the boredom or the frustration self-assessment question was "3" were excluded. As a result, dataset B consisted of 168 trials in total, 55 NB, 70 B/NF and 43 B/F.

### IV. RESULTS

Initially, the subjects' answers to the mid-trials questionnaires were analyzed on the basis of Kendall's tau correlation coefficient, examining correlations between boredom, frustration and the tendency to resign from game playing. Boredom correlated inversely to the subject's willingness to continue playing ( $\tau = -0.784$ ,  $p < 0.001$ ,  $N=195$ ). Inverse correlation was also found between the latter and the player's frustration ( $\tau = -0.208$ ,  $p < 0.001$ ,  $N=195$ ); frustration and boredom were also found to correlate ( $\tau = 0.325$ ,  $p < 0.001$ ). These results support the fact that boredom and frustration are two negative emotional states of great importance in the context of video-games. Their efficient automatic recognition from future game-playing systems could contribute towards ensuring game-playing quality and player satisfaction.

In order to examine whether the moment-based features under consideration can improve automatic recognition of frustration that is induced during HCI, an LDA-based classifier was utilized, trying to solve classification problems related to frustration detection, which were formulated by the annotated data sets described above. Following leave-one-out cross validation, the LDA weights as well as the class centroids were calculated on the basis of the train set, and each test case was classified as belonging to its less distant class, similarly to [13,21]. Classification accuracy was assessed in terms of the correct classification rate (CCR = number of all cases correctly classified / total number of cases). In order to find features with the best discrimination capabilities between emotional classes, a Sequential Backward Search (SBS) [21] feature selection process was employed, using the CCR as the feature selection criterion. SBS was applied on several initial feature sets (some of them described in Table II), as explained in the following.

TABLE II. DESCRIPTION OF FEATURE SETS WHERE SBS WAS APPLIED, CONSISTING OF BOTH GSR AND IBI FEATURES

Feature Set	Features
CONV	All conventional features extracted from GSR and IBI
M	<i>gsr_KrXX</i> and <i>ibi_LgXX</i> features (XX=0-39)
M <sub>mod</sub>	<i>gsr_Kr<sub>mod</sub>XX</i> and <i>ibi_Lg<sub>mod</sub>XX</i> features (XX=0-39)
CM	CONV and <i>gsr_KrXX</i> and <i>ibi_LgXX</i> features (XX=0-39)
CM <sub>mod</sub>	CONV and <i>gsr_Kr<sub>mod</sub>XX</i> and <i>ibi_Lg<sub>mod</sub>XX</i> features (XX=0-39)

A. Frustration Detection with Conventional Features

Using initially only GSR or IBI conventional features as initial feature sets for SBS, max average CCRs of 67.69% (132/195; NF: 102/ 149, F: 30/46) and 78.46% (153/195; NF: 115/149, F: 38/46) were respectively achieved. By fusing the GSR and IBI conventional features, feature set CONV was formed, from which SBS selected features: GSR {Mean, SD, 1<sup>st</sup> Deriv avg, 1<sup>st</sup> Deriv RMS, # of SCRs, Avg SCR Amplitude,  $\delta$ ,  $\delta_{norm}$ ,  $f_d$ ,  $\gamma_{norm\_FL}$ ,  $f_{d\_FL}$ }, IBI {SD, LF/HF, RMSSD, pNN50,  $\gamma_{norm}$ ,  $f_d$ , Mean\_FL, LF/HF\_FL, RMSSD\_FL,  $\delta_{FL}$ ,  $\delta_{norm\_FL}$ }. These features achieved a max CCR of 83.59% in dataset A (Table III). In line with findings of previous works, the joint use of conventional GSR and IBI features was found effective towards automatic frustration detection, yet at a relatively limited accuracy level.

TABLE III. CONFUSION MATRIX AFTER SBS ON CONV FEATURE SET

Annot ated as	Classified as NF	Classified as F	Total	CCR per Class
NF	126	23	149	84.56%
F	9	37	46	80.43%

B. Frustration Detection with Moment-based Features Only

Krawtchouk and Legendre moment-based features were found in [13] the most descriptive transformations of the GSR and IBI modalities respectively. Following this line, SBS was applied in the present study to feature sets *Kgsr* and *Libi*; the first contained the 40 *gsr\_KrXX* features and the second the 40 *ibi\_LgXX* ones. With *Kgsr*, a max CCR of 77.44% (151/195; NF: 123/149, F: 28/46) was achieved in dataset A, whereas *Libi* produced a 68.21% (133/195; NF: 104/149, F: 29/46) CCR. Then, feature set M was formed by fusing *Kgsr* and *Libi*. With this feature set, a max CCR of 80.51% (157/195; NF: 124/149, F: 33/46) was obtained.

Applying similar analysis for the moment-based feature variations, two further feature sets were fed to the SBS, *K<sub>mod</sub>gsr* and *L<sub>mod</sub>ibi*, consisting of all *gsr\_Kr<sub>mod</sub>XX* and *ibi\_Lg<sub>mod</sub>XX* features extracted respectively. *K<sub>mod</sub>gsr* produced a max CCR of 75.90% (148/195; NF: 122/149, F: 26/46) and *L<sub>mod</sub>ibi* achieved 71.79% (140/195; NF: 116/149, F: 24/46). A further initial feature set was formed (*M<sub>mod</sub>*) by fusing the above features, over which the SBS procedure produced a max CCR of 82.56% (161/195; NF: 129/149, F: 32/46). Concluding, by completely replacing conventional features with moment-based ones, frustration detection accuracies close to the initial one (of the CONV feature set) were achieved in dataset A.

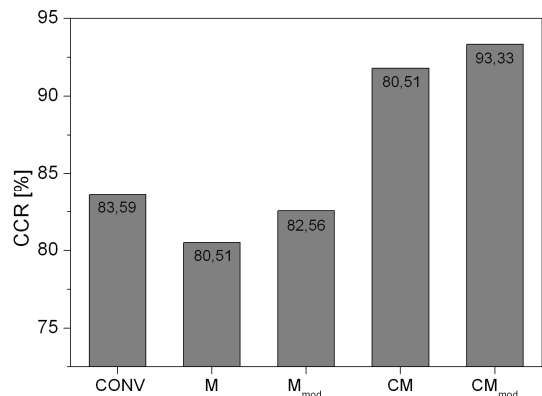


Figure 1. Max average CCRs obtained over Dataset A from the best features selected.

C. Fusion of Conventional and Moment-based Features

SBS was then applied to feature sets built from fusing the conventional features with the moment-based ones. Feature set CM consisted of the conventional features, together with all *gsr\_KrXX* and *ibi\_LgXX* ones. In the CM<sub>mod</sub> feature set, the *gsr\_Kr<sub>mod</sub>XX* and *ibi\_Lg<sub>mod</sub>XX* moment-based feature variations were fused with the conventional features. SBS over CM produced a CCR of 91.79% (179/195; NF: 142/149, F: 37/46), significantly higher (by 8.2%) than the result obtained from CONV. Moreover, SBS over CM<sub>mod</sub> achieved even higher frustration detection accuracy (Table IV); the best model built after SBS contained in this case features: GSR{*gsr\_Kr<sub>mod</sub>XX*; XX=2,4,5,8,9,14,19,20-22,26,27,39}, IBI{Mean, SD, LF/HF, pNN50,  $\gamma_{norm}$ ,  $f_d$ , Mean\_FL,

LF/HF\_FL,  $\delta_{norm\_FL}$ ,  $\gamma_{norm\_FL}$ , *ibi\_Lg<sub>mod</sub>***XX**; **XX**={3-6,11,20,27,36,39}.

TABLE IV. CONFUSION MATRIX OF SBS ON *CM<sub>mod</sub>* (CCR=93.33%; 182/195)

Annotated as	Classified as NF	Classified as F	Total	CCR per Class
NF	140	9	149	84.56%
F	4	42	46	80.43%

As shown in Figure 1, fusing conventional features with moment-based ones significantly increased the accuracy of frustration detection among dataset A. Both *CM* and *CM<sub>mod</sub>* feature sets increased the max average CCR compared to *CONV*; by a maximum 9.74% in the latter case. The significance of this increase in performance was proved by a two-tailed paired t-test ( $p < 0.001$ ). Within the best final feature set (selected from *CM<sub>mod</sub>*), all conventional GSR features were replaced by moment-based ones, indicating the significance of the GSR moment-based feature variations in the context of automatic frustration detection. Such an example is the GSR SD, which although selected from *CONV*, it was discarded from SBS in the *CM<sub>mod</sub>* case and replaced by moment-based GSR features, despite the fact that the specific feature has been found in the past [4] particularly significant towards automatic frustration detection. Regarding the IBI signal, several moment-based features were selected in the final best model built; however they were not capable to totally replace conventional ones. Some of the latter (e.g. pNN50) were kept in the best model, and this underlines their significance towards automatic frustration detection. Nevertheless, it has to be noted that another such feature, RMSSD, was replaced by moment-based features in the best final model built.

D. Joint Automatic Detection of Boredom and Frustration

The effectiveness of moment-based features was assessed also on the basis of a three-class ER problem, towards building a system capable to detect either not-bored, bored, or subjects being bored and frustrated as well. In order to do so, SBS was applied over feature sets *CONV* and *CM<sub>mod</sub>*, in respect of the afore-described dataset B. As shown from Table V, the joint use of moment-based features with conventional ones significantly increased (by 11.91%) the total accuracy of the LDA-based classifier over the given 3-class ER problem.

It has been shown in the past that more-than-two-class ER problems can be effectively split down into simpler binary ones, so as to increase ER efficiency [21]. Following this line, the afore-described original 3-class joint boredom/frustration recognition problem was also split into two binary ones; boredom and frustration detection. Two binary LDA classifiers were used in cascade, the first regarding boredom (LDA-b) and the second regarding frustration recognition (LDA-f). Cases were first classified as B/NB by LDA-b. Then, cases classified as B were fed to LDA-f, which decided whether the subject was also frustrated (B/F) or not (B/NF). Again, two feature set types

were examined, *CONV\_bf* and *CM<sub>mod</sub>\_bf*. For *CONV\_bf*, the best features selected from *CONV* in Section 4.1 and *F\_Set\_C* in [13] were used for the LDA-f and the LDA-b classifiers respectively. For *CM<sub>mod</sub>\_bf*, the best combinations of conventional and moment-based features reported in Section 4.3 and [13] were used for the LDA-f and the LDA-b classifiers respectively.

TABLE V. CONFUSION MATRICES PER FEATURE SET FOR THE 3-CLASS ER PROBLEM

Feature Set	Cases Annotated as	Cases Classified as			Total Cases Nr	CCR	
		NB	B/NF	B/F		Per Class	Total
<i>CONV</i>	NB	41	10	4	55	74.55%	70.83% 119/168
	B/NF	14	44	12	70	62.86%	
	B/F	1	8	34	43	79.07%	
<i>CM<sub>mod</sub></i>	NB	45	9	1	55	81.82%	82.74% 139/168
	B/NF	8	55	7	70	78.57%	
	B/F	1	3	39	43	90.70%	

Following this approach allowed conventional features to achieve an average CCR of 76.19% (128/168) among the 3 classes, significantly increased (by 5.36%) compared to the respective result shown in Table V. Similarly, in the case of *CM<sub>mod</sub>\_bf*, accuracy reached a CCR of 88.69% (149/168), increased by 5.95% compared to Table V. These results further depict the contribution of moment-based features in the domain of automatic ER; automatic multi-class ER systems based on conventional features can be enhanced towards increased efficiency by various techniques proposed in the past (e.g. [21]), and augmenting them with moment-based features can lead to even increased effectiveness.

V. CONCLUSIONS

In this work, experimental evaluation showed that augmenting conventional biosignal features with moment-based ones, significantly enhances the efficiency of binary frustration detection (NF vs. F), which is induced during HCI. Moment-based features were also found effective over a joint frustration and boredom detection ER problem, regarded from a 3-class perspective (NB vs. B/NF vs. B/F). When this problem was split into simpler binary ones, the accuracy of conventional features increased. However, the highest CCR was once more obtained by conventional features fused with moment-based ones.

Biosignal sensors are anticipated to become wireless, smaller and less obtrusive in the future. This will pave the way for future practical HCI systems augmented with biosignals-based automatic affect detection capabilities. Such a case could be an affective game playing system that will be capable to understand in real-time whether negative emotions like frustration have appeared and subsequently adapt, so as to ensure game-playing quality. Similar rationale can be followed in further HCI constructions as well, like e-learning systems etc. It has to be noted however, that in the present study, the real-time monitoring of frustration was not

the immediate target, and off-line processing was applied to biosignals. Nevertheless, the on-line calculation of biosignal features based on the theory of moments can be regarded as feasible in future developments, since the processing power of modern PCs, along with multi-threading techniques already allow the simultaneous real-time extraction of large sets of biosignal features.

The results of the present analysis clearly show that moment-based features are significantly helpful towards enhancing effectiveness in automatic detection of negative emotions like frustration induced during HCI, which is in turn expected to be of great importance towards future affective game-playing systems and other HCI applications as well.

#### REFERENCES

- [1] J. Klein, Y. Moon, R. W. Picard. This computer responds to user frustration: Theory, design, and results, *Interacting with Computers*, Vol. 14, Is. 2, Feb. 2002, Pages 119-140.
- [2] C. Yun, D. Shastri, I. Pavlidis, Z. Deng, 2009. O' game, can you feel my frustration?: improving user's gaming experience via stresscam. *CHI 2009*, 2195-2204.
- [3] R.L. Mandryk, K.M. Inkpen and T.W. Calvert, Using psychophysiological techniques to measure user experience with entertainment technologies, *Behaviour and Information Technology* 25 (2) (2006), pp. 141-158.
- [4] G. N. Yannakakis, H. P. Martinez, and A. Jhala, "Towards Affective Camera Control in Games," *User Modeling and User-Adapted Interaction*, vol. 20, no. 4, pp. 313-340, 2010
- [5] Stephen H. Fairclough, *Fundamentals of physiological computing*, *Interacting with Computers*, Volume 21, Issues 1-2, Special issue: Enactive Interfaces, Jan. 2009, pp. 133-145.
- [6] Gilleade, K.M., Dix, A. (2004). Using frustration in the design of adaptive videogame. Paper presented at the *Advances in Computer Entertainment Technology*.
- [7] Eva Hudlicka. Affective game engines: Motivation and requirements. In *Proceedings of the 4th Int. Conf. on the Foundations of Digital Games*, pages 299-306. ACM, 2009
- [8] N. K. Sivaraman, L. L. Narayana Rao, K. L. Nitin, An Emotional System for Effective and Collaborative e-Learning, *Fourth International Conference on Advances in Computer-Human Interactions, ACHI '11*, 2011
- [9] Mayer, C.; Wimmer, M.; Eggers, M.; Radig, B.; , "Facial Expression Recognition with 3D Deformable Models," *Advances in Computer-Human Interactions, 2009. ACHI '09. Second International Conferences on* , vol., no., pp.26-31, 1-7 Feb. 2009
- [10] Z. Zeng,, M. Pantic, G. I. Roisman, T. S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions", *IEEE T. Pattern Anal. Mach. Intell.* 31(1): 39-58, 2009.
- [11] S. McQuiggan, S. Lee, and J. Lester, Early prediction of student frustration, In *Proc. of the 2nd Intl. Conf. on Affective Computing and Intelligent Interaction*, Portugal, 2007
- [12] Ashish Kapoor, Winslow Burleson, Rosalind W. Picard, Automatic prediction of frustration, *Int. J. Hum.-Comput. Stud*, Volume 65, Issue 8, August 2007, Pages 724-736.
- [13] D. Giakoumis, D. Tzovaras, K. Moustakas, G. Hassapis, "Automatic Recognition of Boredom in Video Games using novel Biosignal Moment-based Features". *Affective Computing, IEEE Transactions on*, Accepted for Publication
- [14] J. Scheirer, R. Fernandez, J. Klein, R. W. Picard, Frustrating the user on purpose: a step toward building an affective computer, *Interacting with Computers*, 14(2),pp.93-118, 2002.
- [15] C.L. Lisetti and F.Nasoz, Using noninvasive wearable computers to recognize human emotions from physiological signals, *EURASIP J. Appl. Signal Process*, 1672-1687, Jan. 2004
- [16] Calvo, R.A.; D'Mello, S., "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications," *IEEE Trans Affective Computing*, 1(1), pp.18-37, Jan. 2010
- [17] M.R. Teague, "Image analysis via the general theory of moments", *J. Opt. Soc. Am.* 70 (8), pp. 920-930, 1980.
- [18] P-T. Yap, R. Paramesran, "Image Analysis by Krawtcouk Moments", *IEEE T. Image Process*, 12 (11) pp. 1367-1377, Nov. 2003.
- [19] T. W. Malone, "What makes computer games fun?," *Byte*, vol. 6, pp. 258-277, 1981
- [20] Scherer KR. "On the nature and function of emotion: A component process approach." In: Scherer KR, Ekman P, eds.- *Approaches to emotion*. Hillsdale, NJ: Erlbaum, 1984:293-318
- [21] J. Kim, and E. Andre, "Emotion recognition based on physiological changes in music listening", *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(12):2067-2083, 2008

# Adaptivity Considerations for Enhancing User-Centric Web Experience

Marios Belk

Department of Computer Science  
University of Cyprus  
Nicosia, Cyprus  
belk@cs.ucy.ac.cy

Panagiotis Germanakos

Institute of Informatics and Telematics (IIT)  
Italian National Research Council (CNR)  
Pisa, Italy  
pgerman@cs.ucy.ac.cy

Panagiotis Zaharias

Department of Management Science & Technology  
Athens University of Economics and Business  
Athens, Greece  
pz@aueb.gr

George Samaras

Department of Computer Science  
University of Cyprus  
Nicosia, Cyprus  
cssamara@cs.ucy.ac.cy

**Abstract** - User Experience (UX) design relates to the creation of models that affect user experiences during interaction with a system, while the incorporation of cognitive factors in the personalization process of Web systems might provide a more user-centric approach. This paper explores the influence specific cognitive factors may have on UX qualities to be used as adaptivity factors for personalizing and improving users' experiences in commercial Web-sites. A user experience evaluation was conducted where 96 students navigated in an existing commercial Web-site for a problem-based task. A user experience measurement was performed so as to assess users' perceptions regarding the pragmatic, hedonic and attraction qualities of the environment. It has been observed that specific cognitive factors have considerable influence on specific qualities of user experience. To this end, such findings are encouraging for further investigation of the possible impact of cognitive factors in terms of enhancing the personalization process of commercial Web-sites so as to achieve better user experience.

**Keywords** - *User-adaptive systems; User Modeling (UM); User Experience (UX); Cognitive Factors*

## I. INTRODUCTION

User-adaptive systems [1] have become progressively popular in the last decade due to the exponential increase of users and availability of digital information, mainly on the World Wide Web (WWW). Since its early days, research on User-adaptive systems focused on the weaknesses presented in traditional "one-size-fits-all" systems [2] that were unable to satisfy the heterogeneous needs and preferences of users.

By User-adaptive systems we mean all computer-mediated systems that are able to (semi) automatically adapt their structure and presentation by learning from data about their users; implicitly (i.e., observing user's interactions) and/or explicitly (i.e., direct input from the user). All User-adaptive systems, from research-oriented to industrial

systems, and from educational hypermedia systems to commercial Web systems, they all share a common goal, i.e., to increase the *functionality* of the system and improve the *users' experiences* by making it personalized.

Research on User-adaptive systems straddles the boundaries of User Modeling (UM) [3] and User Experience (UX) [4]. UM in User-adaptive systems deals with what information is important about the user and how to learn and represent this information, while UX studies the feelings and thoughts of an individual about a product (e.g., interactive system). UX mainly takes the affective consequences on the human side as opposed to the computer side. It focuses on positive emotional outcomes such as joy, fun and pride and deals with hedonic and affective (e.g., surprise, diversion, intimacy) aspects of HCI design and evaluation.

The work presented in this paper lies on previous research [5] that has shown that individual traits (i.e., specific cognitive factors) may have significant impact in the adaptation and personalization process of User-adaptive systems. Furthermore, current advances in User Experience (UX) reveal that there is still enough space for research by combining UM and UX strategies for providing a more user-centric approach in User-adaptive systems.

In this respect, this work is an attempt to study a possible association between specific cognitive factors of a cognitive-based user model [5] and UX qualities of a valid UX measurement (AttrakDiff) [12]. The main aim is to identify possible relationships between these cognitive factors and UX qualities that will generate new research possibilities; personalizing commercial Web-sites based on cognitive factors in order to improve users' experiences (focused on hedonic and affective aspects) during interaction with a product.

The remainder of the paper is organized as follows: Sections 2 and 3 briefly discuss the research areas related with this work; User Modeling (UM) in User-adaptive system and User Experience (UX), respectively. Section 4



formulates the basis for the investigation of the relationship between cognitive factors and UX qualities. Section 5 presents the experimental methodology of a study with 96 participants, followed by a discussion on its results. The paper concludes with some future prospects.

## II. USER MODELING IN USER-ADAPTIVE SYSTEMS

Adaptation decision in User-adaptive systems is based on taking into account any vital information about the user, represented in the user model, in order to provide adaptation effects (i.e., the same system can look different to users with different models). The data kept in the user model can be distinguished (according to [6]) to *user data*, *usage data*, and *environment data*. User data comprise various characteristics about the user, usage data comprise data about user's interactions with the system that are utilized to infer knowledge about the user, and environment data comprise all aspects of the user's environment (i.e., context, device's or network's characteristics).

The work presented in this paper focuses on modeling *user data*. Among the five most popular user features (i.e., user's knowledge, interests, goals, background, and individual traits) of user data applied in User-adaptive systems [3], this study focuses on users' individual traits (i.e., cognitive factors) for modeling the user.

The user's individual traits are features that define the user as an individual. Examples are personality factors, cognitive factors and learning styles. Unlike other features, that are extracted through interviews or based on user's interactions, individual traits are traditionally extracted using psychometric tests. Individual traits are stable user features that might change only over a long period of time or might not change at all, in contrast with other user features, such as user's goals, knowledge, interests that are rather dynamic features and change frequently over time.

Recently, a considerable amount of research efforts have been undertaken focusing on modeling and utilizing cognitive factors for personalization in User-adaptive systems. Several User-adaptive systems [5][28][29][30] have distinguished users based on their cognitive styles and learning styles and provided different adaptation effects accordingly. In a study, Germanakos et al. [5] have distinguished imager and verbal users, and wholist users and analyst users based on Riding's Cognitive Style Analysis [16]. Each user was provided with adaptive presentation of content and different navigation organization. In a similar approach, Triantafillou et al. [28] distinguished field-dependent and field-independent users based on Witkin et al. [18] and provided different navigation organization, amount of user control, and navigation support tools for these groups. Results in both studies indicate that cognitive styles have significant impact in the adaptation and personalization process of Web environments by increasing usability and user satisfaction during navigation and learning performance.

Such findings suggest that individual traits are important user features to take into account in the personalization process of a User-adaptive system.

## III. USER EXPERIENCE (UX)

ISO 9241-210 [7] defines User Experience (UX) as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service". UX is dynamic, because it changes over time as the circumstances change [8]. Being a multi-dimensional and complicated area a universal definition has not been agreed to date. Nevertheless, most of the definitions given to UX [9, 10] agree that UX focuses on the hedonic and affective aspects of HCI, but it also includes a person's perceptions of the practical aspects such as utility, ease of use and efficiency of a system.

Effective HCI design and evaluation involves two important qualities: i) usability (i.e., traditional HCI), and ii) hedonic, beauty and affective [10]. Based on Jordan [11], the latter complements traditional HCI qualities (i.e., pragmatic) by suggesting a fixed hierarchical structure of qualities that contribute to positive experience. That is, a product has to provide functional and usability qualities before hedonic aspects can take effect. In contrast to Jordan, Karapanos et al. [8] assume the importance of these different qualities to vary with several contextual factors, i.e., individual differences, type of product, situation the product is used in, and change of experience over time.

Regarding UX evaluation, one of the most influential models is the one proposed by Hassenzahl [12]; according to this model each interactive product has a pragmatic (related to usability) and hedonic quality that contribute to the UX. Based on this model a well-known and widely used measurement instrument has been developed, the AttrakDiff [12], which has been employed in our empirical study (version AttrakDiff2). It is composed of four main constructs with seven anchor scales (total 28 items). Within each item a word-pair spans a scale between two extremes. The scales consist of seven stages (-3,-2,-1,0,1,2,3) between the word-pairs. The oppositional word-pairs consist of two conflictive adjectives like "bad" – "good", or "technical" – "human".

The constructs are [12][13]: Pragmatic Quality (PQ), which is related to traditional usability issues (such as, effectiveness, efficiency, learnability, etc.); Hedonic Quality Stimulation (HQ-S), which is about personal growth of the user and the need to improve personal skills and knowledge; Hedonic Quality Identification (HQ-I), which focuses on the human need to be perceived by others in a particular way; and Attraction (ATT), which is about the global appeal of an interactive product.

## IV. RELATIONSHIP ANALYSIS BETWEEN COGNITIVE FACTORS AND UX QUALITIES

This section presents specific cognitive factors that could influence UX qualities (based on the AttrakDiff analysis) in User-adaptive systems. The goal is to initially investigate and formulate a cognitive-based user model for User-adaptive systems in relation to the UX perspective (Figure 1).

A. Cognitive Styles

Cognitive styles represent the particular set of strengths and preferences that an individual or group of people have in how they take in and process information. By taking into account these preferences and defining specific strategies, empirical research has shown that cognitive styles correlate with performance in a Web-based environment [14][21]. Cognitive styles have been defined by Messick as “consistent individual differences in preferred ways of organizing and processing information and experience, a construct that is different than learning style” [15].

Regarding the hypermedia information space, amongst the numerous proposed theories of individual style, a selection of the most appropriate and technologically feasible cognitive (and learning) styles (those that can be projected on the processes of selection and presentation of Web-content and the tailoring of navigational tools) has been studied, such as Riding’s Cognitive Style Analysis (CSA) (Verbal-Imagery, and Wholistic-Analytic) [16], Felder/Silverman Index of Learning Styles (ILS) (4 scales: Active vs. Reflective, Sensing vs. Intuitive, Visual vs. Verbal, and Global vs. Sequential) [17], Witkin’s Field-Dependent, and Field-Independent [18], and Kolb’s Learning Styles (Converger, Diverger, Accommodator, and Assimilator) [19], in order to identify how users transform information into knowledge (constructing new cognitive frames).

TABLE I. RIDING COGNITIVE STYLE SCALE

CSA Scale	Typology	Description	Web Implications
Imagery-Verbal	Imager	Represents information in mental pictures	Prefers graphic, pictorial/visual representation
Imagery-Verbal	Intermediate	No specific preference	Combination of graphics and text
Imagery-Verbal	Verbal	Represents information verbally	Prefers material in text/auditory form
Wholistic-Analytic	Wholist	Organizes information as whole	Needs more guidance, serial navigation approach
Wholistic-Analytic	Intermediate	No specific preference	Combination of scattered and serial navigation
Wholistic-Analytic	Analyst	Organizes information in parts	Independent and scattered navigation

Riding and Cheema’s CSA [20] has been used as a reference theory of cognitive style in previous research [21] due to the fact that the two independent scales (Table 1) of the CSA (Imagery-Verbal, and Wholistic-Analytic) correspond at a considerable extend to the structure of hypermedia (i.e., Web) environments. A personalized environment that is supported by an automated mechanism can be altered mainly at the levels of content selection and hypermedia structure; the content is essentially either visual or verbal (or auditory), while the manipulation of links can lead to a more analytic and segmented structure, or to a more holistic and cohesive environment. These are actually the

differences in the preferences of individuals that belong to each dimension of the CSA scale [22].

This study utilizes Riding’s Cognitive Style Analysis (CSA) [16] because its implications can be mapped on the information space more precisely, since they consist of distinct scales that respond directly to different aspects of the Web space. The CSA implications (Table 1) are quite clear in terms of hypermedia design (visual/verbal content presentation and wholistic/analytic pattern of navigation).

In this respect, the following research assumptions could be formulated: The Imagery-Verbal factor, which mainly influences the presentation of content (visual/textual), might primarily affect the Attraction (ATT) and Hedonic Quality Stimulation (HQ-S) constructs that relate to the overall appeal of a Web-site and stimulation and attention through inspiring and supportive content as well as interaction/presentation techniques. Furthermore, the Wholist-Analyst factor might primarily affect Pragmatic Quality (PQ) that is related to traditional usability issues (i.e., instrumental efficiency and effectiveness) because this factor has high impact on the navigation and instrumental functionality (e.g., navigation support, supportive tools) of a Web-site.

B. Working Memory Span

The concept of Working Memory Span (WMS) [23] also fits very well into our rationale [21] of personalizing Web content on the basis of users’ cognitive abilities and preferences. “The term working memory refers to a brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language comprehension, learning, and reasoning” [24]; Baddeley also refers to individual differences in the WMS (digit) of the population, thus providing a very good argument for using this concept as a personalization factor.

We are mainly interested in the notion of the WMS; since it can be measured and the implications on information processing are rather clear (Table 2). Each WMS instance (i.e., low/medium/high), indicating the working memory capacity of a person, has implications on the navigation, quantity of content and aesthetics of a Web environment. Due to the visual form of presentation in the Web, we have focused especially on the measurement of visual WMS [25] in terms of psychometrics.

TABLE II. WORKING MEMORY SPAN SCALE

WMS	Description	Web Implications
Low	Low working memory capacity	Needs more guidance in Web-sites, navigation support tools, less content, emphasize content
Medium	Medium working memory capacity	Prefers navigation support tools, emphasize content
High	High working memory capacity	More complex structure of Web-sites, more content

The idea of exploring the role of working memory in hypermedia environments has indeed generated research. DeStefano and LeFevre [26] reviewed 38 studies that address mainly the issue of cognitive load in hypertext

reading, and working memory is often considered as an individual factor of significant importance, even at the level of explaining differences in performance. Lee and Tedder [27] examine the role of working memory in different computer texts, and their results show that low WMS learners do not perform equally well in hypertext environments.

The research assumption in this case is that WMS might significantly affect Pragmatic Quality (PQ), Hedonic Quality Stimulation (HQ-S) and Attraction (ATT), since WMS has implications on the navigation, quantity of content and aesthetic appeal of a Web-site.

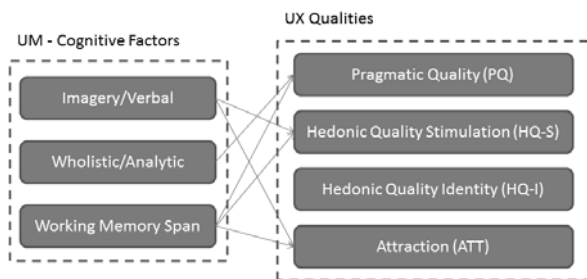


Figure 1. Relationship Between Cognitive Factors and UX Qualities

Concluding, Hedonic Quality Identity (HQ-I) might not significantly affect any of the cognitive factors in the current context since it is essentially intended for evaluation of hedonic qualities of products, rather than software (i.e., Web-based environments). Thus, the HQ-I construct is not included in the current analysis.

V. STUDY

In this section, we describe a preliminary study that aims to evaluate the experiences users had in a commercial Web-site based on their cognitive factors and support the aforementioned possible relationship between UX qualities. A UX measurement (AttrakDiff2) was performed so as to assess their perceptions regarding the pragmatic and hedonic qualities of the environment.

A. Methodology and Sampling

The study was carried out at the University of Cyprus during the whole month of March 2011. Our sample included 96 students of the Computer Science department. The participation was voluntarily. All participants' ages varied from 18 to 21, with a mean age of 19. All users accessed a commercial Web-site using personal computers located at the laboratories of the university, divided in groups of approximately 20 participants. Each session lasted about 40 minutes; 20 minutes was required for the user modeling process, while the remaining time was devoted to navigate in an existing commercial Web-site and evaluate their experiences using AttrakDiff2.

During the user modeling process, students provided their demographic characteristics (i.e., name, age, education, etc.) and performed a number of interactive tests using attention and cognitive processing efficiency grabbing

psychometric tools [31] in order to quantify their cognitive characteristics.

Furthermore, the students were asked to navigate in a replica of the official Web-site of HTC Corp. (www.htc.com, derived on March 1, 2011) that was developed for the purpose of this experiment. The Web-site's content was about a series of mobile phones; general description, technical specifications and additional information were available for each model.

The students were asked to fulfill three tasks; they had to find the necessary information to answer three sequential multiple choice questions that were given to them while navigating and which were referring to a particular type of mobile phone. There was certainly only one correct answer that was possible to be found relatively easy, in the sense that the students were not required to have hardware related knowledge or understanding.

As soon as they completed the three tasks, they were presented with an online version of AttrakDiff2 to express their opinion regarding the hedonic and pragmatic qualities of the environment they just navigated.

B. Results

In order to assess the significance and possible impact cognitive factors may have on UX qualities, a comparison has been performed between the cognitive factor's instances per UX construct.

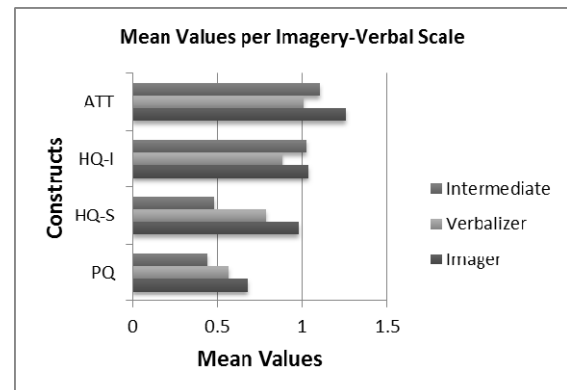


Figure 2. Mean UX Values per Imagery-Verbal Scale

Figure 2 illustrates the mean values of the UX dimensions per Verbal-Imagery scale. According to the results, we observe considerable deviation between the instances in Hedonic Quality Stimulation (HQ-S) and Attraction (ATT). In HQ-S, mean scores of Imagery were 0.98, Verbalizers 0.79 and Intermediates 0.48, and in ATT, mean scores of Imagery were 1.26, Verbalizers 1.01 and Intermediates 1.1, indicating that users based on this cognitive factor perceived differently the hedonic quality (stimulation) and overall appeal (attraction) of the environment. In addition, based on an empirical observation of the environment, it can be easily revealed that the environment is rich with graphical/visual representations.

To this end, the Imagery-Verbal factor primarily influences the HQ-S and ATT constructs and that Imagery

find the environment more stimulating and attractive since the environment contains a lot of graphical representations.

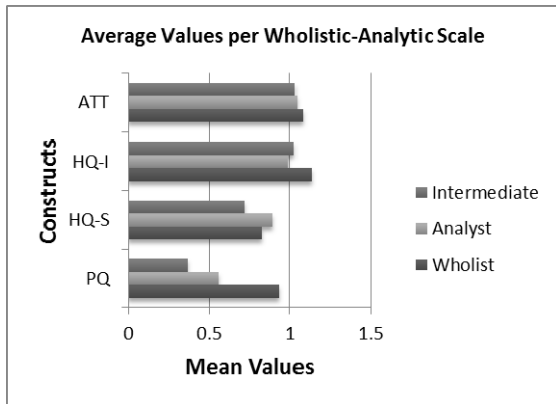


Figure 3. Mean UX Values per Wholistic-Analytic Scale

Figure 3 illustrates the mean values of the UX dimensions per Wholistic-Analytic scale. According to the results, we observe considerable deviation between the instances in Pragmatic Quality (PQ). In PQ, mean scores of Wholists were 0.93, Analysts 0.79 and Intermediates 0.36 indicating that users based on this cognitive factor perceived differently the pragmatic quality (i.e., usability) of the environment. Regarding the Attraction (ATT), marginal deviation has been observed, with Wholists having mean scores of 1.08, Analysts 1.05 and Intermediates 1.03, indicating that the Wholistic-Analytic factor does not significantly influence ATT. This is in accordance with the implications the Wholistic-Analytic factor has on the Web space; it primarily influences usability of a Web-site (i.e., different navigation, support tools for different user models).

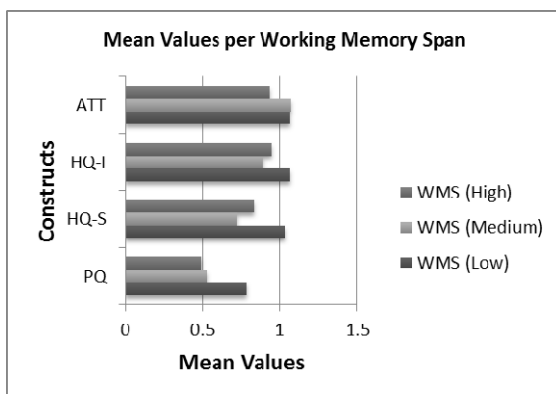


Figure 4. Mean UX Values per Working Memory Span

Figure 4 illustrates the mean values of the UX dimensions per Working Memory Span. According to the results, we observe considerable deviation between the instances in Hedonic Quality Stimulation (HQ-S) and Pragmatic Quality (PQ). In HQ-S, users with low working memory span (WMS) had mean scores of 1.03, with medium WMS 0.72 and with high WMS 0.84, and in PQ, mean scores of low WMS users were 0.78, medium WMS users

0.53 and high WMS users 0.49, indicating that users based on these cognitive factors perceived differently the hedonic quality (stimulation) and usability of the environment. In Attraction (ATT), minor deviation of ratings has been observed, indicating that this construct is not primarily influenced by this cognitive factor.

In conclusion, the WMS factor primarily influences the HQ-S and PQ constructs and that users with low WMS find the environment more stimulating and “usable” since the environment contains supportive content and a lot of graphical indications.

## VI. CONCLUSION AND FUTURE WORK

In conclusion, this empirical work focused on the influence specific cognitive factors may have on UX constructs. A UX measurement was performed so as to assess users’ perceptions regarding the pragmatic and hedonic qualities of a commercial Web-site.

Initial findings indicate that incorporating specific cognitive factors in the personalization process of commercial Web-sites may affect users’ experiences. Results reveal that the Imagery-Verbal factor primarily affects Hedonic Quality Stimulation (HQ-S) and Attraction (ATT), the Wholistic-Analytic factor primarily affects Pragmatic Quality (PQ), and the WMS factor primarily affects Hedonic Quality (HQ-S) and Pragmatic Quality (PQ). In addition, results indicate that Hedonic Quality Identity (HQ-I) does not primarily affect any of the cognitive factors in the current context, which is in accordance with theory [12], since HQ-I is essentially intended for evaluation of hedonic qualities of products, rather than software (i.e., Web environments).

The relevant research is in its infancy and further empirical studies are needed to investigate UX issues in such context. A future research prospect is to employ personalization methods [5] in order to assess the impact cognitive factors may have in the personalization process of commercial Web-sites, in terms of amplifying users’ experiences during navigation session. In this respect, an experimental study with real users will be conducted in the future (similar experimental approach to Germanakos et al. [5]), where specific adaptation and personalization techniques will be employed in the same commercial Web-site used in this study. The main aim is to adapt and personalize the original environment based on each individual’s cognitive factors in such a way (e.g., remove some graphical representations of the Web content for Verbal users or enhance navigation support for Wholist users) to achieve positive experience. In order to assess the approach, users will evaluate the original and personalized version of the commercial Web-site utilizing AttrakDiff2. A statistical analysis will be performed in order to compare the user experience evaluations between the two different environments. Based on the influence specific cognitive instances may have on UX qualities, we assume that the personalized version will lead to better emotional outcomes and that the UX qualities will have better ratings than the original version.

Even though the evaluation of this concept in the eCommerce domain is encouraging for our work, there is

still a lot room of investigation in order to shed light on this complex and dynamic research area. Main goal is to initiate and drive this research to a concrete cognitive factors' framework that can be used in any hypermedia system proposing a new set of design guidelines for the enhancement of one-to-one Web services' delivery.

#### ACKNOWLEDGMENT

This work is co-funded by the EU project CONET (INFSO-ICT-224053) and by the project smarTag (within the framework of University of Cyprus internal funded projects).

#### REFERENCES

- [1] M. Schneider-Hufschmidt, T. Kühme and U. Malinowski, *Adaptive user interfaces: Principles and practice*. Elsevier Science Inc., New York, USA, 1993.
- [2] P. Brusilovsky and M.T. Maybury, "From adaptive hypermedia to the adaptive web," *Communications of the ACM*, vol. 45, no. 5, pp. 30-33, 2002.
- [3] P. Brusilovsky and E. Millán, "User models for adaptive hypermedia and adaptive educational systems," In *The Adaptive Web*, vol. 4321, P. Brusilovsky, A. Kobsa and W. Nejdl, Eds. Lecture Notes in Computer Science, Springer, Berlin, 2007, pp. 3-53.
- [4] M. Hassenzahl and N. Tractinsky, "User experience - a research agenda," *Behaviour & Information Technology*, vol. 25, no. 2, pp. 91-97, 2006.
- [5] P. Germanakos, N. Tsianos, Z. Lekkas, C. Mourlas, M. Belk and G. Samaras, "Towards an Adaptive and Personalized Web Interaction using Human Factors," In *Advances in Semantic Media Adaptation and Personalization*, vol. 2, M. C. Angelides, P. Mylonas and M. Wallace, Eds. Auerbach Publications, 2009, pp. 247-281.
- [6] A. Kobsa, J. Koenemann and W. Pohl, "Personalized hypermedia presentation techniques for improving online customer relationships," Technical report, no. 66, German National Research Center for Information Technology, St. Augustin, Germany, 1999.
- [7] ISO FDIS 9241-210:2009, *Ergonomics of human system interaction - Part 210: Human-centered design for interactive systems (formerly known as 13407)*. International Organization for Standardization (ISO), Switzerland, 2009.
- [8] E. Karapanos, M. Hassenzahl and J. Martens, "User experience over time," *Extended Abstracts on Human Factors in Computing Systems (CHI'08)*, 2008, pp. 3561-3566.
- [9] J. Forlizzi and K. Battarbee, "Understanding experience in interactive systems," *Proc. on Designing interactive systems: processes, practices, methods, and techniques (DIS'04)*, ACM, 2004, pp. 261-268.
- [10] M. Hassenzahl and N. Tractinsky, "User experience - a research agenda," *Behaviour & Information Technology*, vol. 25, no. 2, 2006, pp. 91-97.
- [11] P.W. Jordan, *Designing Pleasurable Products: An Introduction to New Human Factors*. Taylor & Francis Group, 2000.
- [12] M. Hassenzahl, "The interplay of beauty, goodness, and usability in interactive products," *Human-Computer Interaction*, vol. 19, 2004, pp. 319-349.
- [13] M. Schrepp, T. Held and B. Laugwitz, "The influence of hedonic quality on the attractiveness of user interfaces of business management software," *Interacting with Computers*, vol. 18, no. 5, 2006, pp. 1055-1069.
- [14] K. H. Wang, T. H. Wang, W. L. Wang and S. C. Huang, "Learning styles and formative assessment strategy: enhancing student achievement in Web-based learning," *Journal of Computer Assisted Learning*, vol. 22, no. 3, 2006, pp. 207-217.
- [15] E. Sadler-Smith, "The relationship between learning style and cognitive style," *Personality and Individual Differences*, vol. 30, no. 4, 2001, pp. 609-616.
- [16] R. Riding, *Cognitive Style Analysis - Research Administration*. Learning and Training Technology, 2001.
- [17] R. Felder and L. Silverman, "Learning and Teaching Styles in Engineering Education," *Engineering Education*, vol. 78, 1988, pp. 674-681.
- [18] H. A. Witkin, C. A. Moore, D. R. Goodenough and P. W. Cox, "Field-dependent and field-independent cognitive styles and their educational implications," *Review of Educational Research*, vol. 47, 1977, pp 1-64.
- [19] A. Kolb and D. Kolb, *The Kolb Learning Style Inventory-Version 3.1*, Technical Specifications, Experience Based Learning Systems, 2005.
- [20] R. Riding and I. Cheema, "Cognitive styles - An overview and integration," *Educational Psychology*, vol. 11, no. 3-4, 1991, pp. 193-215.
- [21] N. Tsianos, P. Germanakos, Z. Lekkas and C. Mourlas, "An Assessment of Human Factors in Adaptive Hypermedia Environments," In *Intelligent User Interfaces: Adaptation and Personalization Systems and Technologies*, C. Mourlas and P. Germanakos, Eds. IGI Global, 2009, pp. 1-34.
- [22] E. Sadler-Smith and R. Riding, "Cognitive style and instructional preferences," *Instructional Science*, vol. 27, no 5, 1999, pp. 355-371.
- [23] A. D. Baddeley, "The concept of working memory: A view of its current state and probable future development," *Cognition*, vol. 10, no. 1-3, 1981, pp. 17-23.
- [24] A. D. Baddeley, "Working Memory," *Science*, vol. 255, 1992, pp. 556-559.
- [25] R. H. Loggie, G. M. Zucco, A. D. Baddeley, "Interference with visual short-term memory," *Acta Psychologica*, vol. 75, no. 1, 1990, pp. 55-74.
- [26] D. DeStefano and J. Lefevre, "Cognitive load in hypertext reading: A review," *Computers in Human Behavior*, vol. 23, no. 3, 2007, pp. 1616-1641.
- [27] M. J. Lee and M. C. Tedder, "The effects of three different computer texts on readers' recall: based on working memory capacity," *Computers in Human Behavior*, vol. 19, no. 6, 2003, pp. 767-783.
- [28] E. Triantafyllou, A. Pomportsis, S. Demetriadis and E. Georgiadou, "The value of adaptivity based on cognitive style: an empirical study," *British Journal of Educational Technology*, vol. 35, 2004, pp. 95-106.
- [29] S. Graf and Kinshuk, "Advanced Adaptivity in Learning Management Systems by Considering Learning Styles," *Proc. of the International Workshop on Social and Personal Computing for Web-Supported Learning Communities (SPeL'09)*, 2009, pp. 235-238.
- [30] K. A. Papanikolaou, M. Grigoriadou, H. Kornilakis and G. D. Magoulas, "Personalising the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE," *User-Modeling and User-Adapted Interaction*, vol. 13, no. 3, 2003, pp. 213-267.
- [31] AdaptiveWeb System, <http://adaptiveweb.cs.ucy.ac.cy>, March, 2011.

# Modality Preferences of Different User Groups

Benjamin Weiss, Sebastian Möller and Matthias Schulz

*Quality and Usability Lab*

*Dt. Telekom Laboratories, Technische Universität*

*Berlin, Germany*

[BWeiss | Sebastian.Moeller | Matthias-Schulz] @tu-berlin.de

**Abstract**—In order to examine user group differences in modality preferences, participants of either gender and two age groups have been asked to rate their experience after interacting with a smart-home system offering unimodal and multimodal input possibilities (voice, free-hand gesture, smartphone touch screen). Effects for gender, but not for age (younger and older adults) have been found for modality preferences. Women prefer touch and voice over gesture for many scales assessed, whereas men do not show this pattern consistently. Instead, they prefer gesture over voice for hedonic quality scales. Comparable results are obtained for technological expertise assessed individually. This interrelation of gender and expertise could not be solved and is discussed along with consequences of the results obtained.

*Keywords*-multimodal dialog system; evaluation; user factors.

## I. INTRODUCTION

Current evaluations of multimodal interfaces already take into account user groups: Differences in users' interactive and rating behavior is analyzed regarding e.g., gender, age, user's experience with a system, etc. Unfortunately, the attitudes and expectations people have towards such systems are not well known yet [1]. Even more, expectations concerning novel multimodal application seem not to be that relevant for the actual user experience [1][2][3]. Modality preference and selection are dependent on task and efficiency [4], but general user expertise [5] also has to be taken into account.

Age and gender effects, for instance, are rarely examined together, with [3][6] as notable exceptions. While in most studies gender is balanced but not looked into further, studies on modality preferences are often limited to younger adults (e.g., [1][7]). Studies including older adults are mostly focusing on assisting technologies to support independent living [8][9][10], but age does not necessarily limit the number of products used [11]. For example, home entertainment and control is one of the major application domains for HCI and also is in the focus of this paper.

Exploring strategies for including older users, multimodality and touch were found to be more suitable than speech and motion control [6]. Furthermore, "older participants used the flexibility offered by the multiple input modalities to a lesser extent than younger users did" [6].

Comparing pointing times on a graphical user interface (GUI) using a mouse or touch-panel no significant difference

between younger, middle-age and older adults was found for touch in contrast to mouse control [12]. The authors conclude that touch interfaces should be pursued to make information technology accessible to older adults.

Experience, although an established feature [13], is typically not a dimension to separate user groups in the field of multimodal systems. Multimodal interfaces are typically innovative and therefore performing evaluation experiments to compare trained versus novice users does not seem to be mandatory. Instead, general technological affinity is assessed in order to analyze this factor.

The aim of this paper is to have a closer look at the interaction between age and gender, as especially for age effects on rating behavior can be expected on the basis of the literature referred to, i.e. overall positive results for older adults [3][6] and gesture preference for younger adults [3]. But, we also want to look into other user differences and their interaction with age or gender. For this purpose, a small battery of assessments has been conducted in order to assess various aspects of technological affinity. For the domain of home entertainment and control user modality preferences (speech, 3d gesture, touch) are analyzed in order to find relevant user attributes to correlate and explain user modality preferences. After presenting the system used, we describe the experimental design, and results of the assessments, as well as the user ratings of the multimodal interaction session and the comparison of the ratings of the unimodal interaction sessions.

## II. MULTIMODAL SYSTEM

For the experimental study, a smart-home system was used offering sequential use of voice, smartphone-based input (touch) and three-dimensional free-hand gesture control (gesture). This system is set up inside a fully functional living room. Possible interactions include the control of lamps and blinds, the TV, an IP-radio, an electronic program guide (EPG), video recorder, and a hi-fi system. Furthermore, the system offers an archive for music and supports the generation of playlists. The TV program, available radio stations, lists of recorded movies, an overview of the users' music (sorted by album, artist, etc.) or the playlist are displayed on the TV screen (cf. Figure 1). Those lists are also displayed on the smartphone to allow touch input for

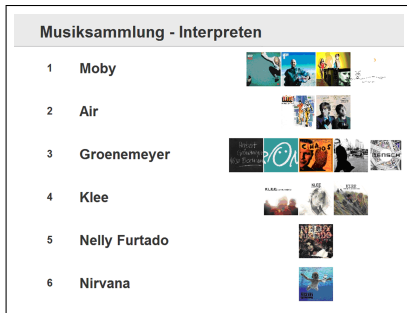


Figure 1. Screen shot of information displayed on the TV screen.

the selection of list entries and the execution of subsequent actions, such as recording a movie or deleting a song from the playlist (cf. Figure 2).

A German male voice was chosen for the TTS system. Thus, three different output channels are employed (TTS, GUI on the touch screen, and lists on the TV screen), in some cases offering complementary or redundant information in parallel. In order to keep input accuracy comparably high for all modalities, speech recognition was replaced by a transcribing wizard. Participants were told that there was a speech recognizer in place, and a lapel microphone was used to further strengthen this impression.

A simple graphical user interface was developed and implemented on an Apple iPhone 3GS, which communicated via wireless LAN with the smart-home system. To control the lamps, blinds, radio and TV the corresponding button on the main screen had to be pressed with a fingertip. This opens a list of options available for the respective device. Further buttons open a music archive, a music playlist, or an overview of recorded movies. List navigation was possible via scrolling (slide finger across screen) and selecting (touching an entry).

A camera-based gesture recognition for simple and often used gestures (TV, radio, lamps, blinds) was simulated by placing two cameras in front of the participants at a distance of approximately two meters, and below the TV screen. The actual recognition was done by the wizard who could monitor the participant via the cameras and enter the recognition result as attribute-value pairs (e. g. [device:blinds; action:down]) into the system. Each fourth of the participants was presented a system with either perfect recognition rate (due to the wizard), reduced speech recognition (10% error rate), reduced gesture recognition (10%) or both (gender and age group balanced).

A set of five three-dimensional gestures was used in this experiment (see Table I). By pointing towards a device with the hand this device is selected. The same gesture could thus be used to initiate the same effect for different devices. Reusing the same concept for different system controls reduces the gesture set considerably. For more detailed information please refer to [14].



(a) Main Screen (b) TV control (c) EPG screen

Figure 2. Screen shots of the smartphone display.

Table I  
GESTURE-COMMAND MAPPING.

Gesture	Command	Device
Swing up	Volume up	TV, Radio
	Brighter	Lamps
	Open	Blinds
Swing down	Volume down	TV, Radio
	Dim	Lamps
	Close	Blinds
Point forward	Turn on/off	TV, Radio
Swing to the right	Next channel	TV, Radio
	Stop	Blinds
Swing to the left	Previous channel	TV, Radio

### III. EXPERIMENT

#### A. Participants

17 young and 17 older adults were asked to participate in the study. For the analysis, data from two subjects (one older male and one younger female adult) has been excluded, as one (younger adult) immediately recognized the Wizard-of-Oz scenario, whereas another (older adult) experienced an unstable system. This results in a group of 16 younger participants (20–29 years, M=24, SD=2.7, 8 female), who have been recruited at the university campus. The 16 older participants (51-67 years, M=59, SD=4.6, 9 female) were selected to also represent the target group of the home entertainment and device control system and thus did not exhibit physical or cognitive disabilities, which would result in special technical requirements. Therefore, they were recruited via notices placed, e.g., in supermarkets. All subjects were paid for their participation. None of the participants was familiar with the system used in this study.

#### B. Procedure

The experiment was split into four parts:

- A: Judgment of the system output (passive scenario)
- B: Judgment of the unimodal input (3 interactive scenarios)
- C: Judgment of the multimodal input (interactive scenario)
- D: Battery of user related assessments

In the first part (Part A), participants were asked to rate each of the three different output channels (TTS, touch screen and TV screen) after the presentation of a series of three to seven examples of one output channel. According to [15], it is sufficient to show a web page for less than one second to judge its aesthetics. Thus, each interface was presented only very shortly to the participants.

In the second part (Part B), the participants were guided through three identical task-based interactions, each time using a different input (touch, voice and gesture). The tasks were short, simple, and closely defined, such as “Lower the blinds and stop them midway” or “Turn on the radio and switch to the next station”. This part was used to collect judgments for each input modality and to train the participants in the use of the modalities and the system. The sequence of output and input in Part A and B followed a full Latin square design to counterbalance order effects.

In the Part C, the user was guided by four tasks displayed one at a time on the screen in front of them. This time participants could choose freely which modality they wanted to use and change the modality whenever they felt like it. The first task consisted of all the interactions that had been conducted in Part B, but in this part the subtasks were less precisely defined (e.g., “Find a radio station you like”). The second and third task asked the participants to do something they had not done before, such as programming a movie or adding songs to their playlist. These tasks could not be solved via gestural interaction. As participants were not explicitly informed about this, some tried nevertheless. The fourth task was open; users were asked to “play” with the system, again try something they had not done yet or use a modality they had not used often.

In the final part (D), each participant had to perform the Digit-Span test [16] to assess memory capacity as control variable and fill out questionnaires assessing technological affinity [17], ICT experience/attitude in order to assess user features apart from age and gender that are potentially related to modality preferences.

C. Assessments

All participants were asked for their ratings of the three output channels (Part A), the three unimodal input channels (Part B) and the multimodal interface (Part C) on the AttrakDiff questionnaire [18], resulting in seven questionnaires filled in per participant (3,3,1). The AttrakDiff questionnaire contains antonym pairs rated on a 7-point scale ([-3,+3]), yielding the subscales *Attractiveness* (ATT), *Pragmatic Qualities* (PQ), *Hedonic Quality – Stimulation* (HQS) and *Hedonic Quality – Identity* (HQI).

According to [19], overall *Attractiveness* (i. e., valence, beauty) is the result of a simple linear combination of PQ (i.e., simple and functional), HQS and HQI. Of the hedonic qualities, *Identity* describes how well a user identifies with the product. *Stimulation* indicates the extent to which a

Table II  
SIGNIFICANT RESULTS FOR ASSESSMENTS OF USER CHARACTERISTICS.  
F-VALUES ( $F_{(1,28)}$ ) AND SIGNIFICANCE LEVEL (ASTERISK).

Data	gender	age	gender:age
Digit-Span value	—	—	—
Technical Expertise (TA)	F=7.29*	F=9.88**	—
Positive Technological Consequences (TA)	—	—	—
Negative Technological Consequences (TA)	—	—	—
Anxiety (ICT)	—	—	F=13.54***
Gadget Loving (ICT)	—	—	—
Training Need (ICT)	—	F=5.62*	—

product supports the needs to develop and move forward by offering novel, interesting and stimulating functions, contents, interactions and styles of presentation.

IV. RESULTS

The reduced recognition rates for some participants did not result in significant rating differences on any ratings scale for any modality condition ( $\alpha = .05$ ) and can thus be neglected in the following.

User variables assessed in Block D are checked for cross-correlations: The following subscales from both questionnaires *technical affinity* (TA) and *ICT attitude/experience* (ICT) have been excluded from analysis, as they seem to assess related constructs due to significant product-moment correlations ( $\alpha = .05, p > .35$ ) with other subscales: *Fascination* (TA) (correlates with *Expertise* (TA) and *Gadget Loving* (ICT)); *Exploratory Behavior* (ICT) (correlates with *Anxiety* (ICT)); *Design Oriented* (ICT) (correlates with *Need For Training* (ICT)); *Riskiness* (ICT) (correlates with *Assumed Negative Consequences* (TA)).

Then, the participants were divided into two groups of age, and gender, respectively. The resulting four groups are tested for differences in the remaining user specific assessments (TA, ICT, Digit-Span test). Age and gender give significant results for some scales assessed (see Table II).

Both age groups do not show any difference in their memorizing abilities. The older adults recruited can be considered as belonging to a possible target group of our multi-modal test system, as they do not exhibit discrepancies in their cognitive abilities and obviously are not physically disabled and thus are not in need of assistive technology.

Self-reported *Expertise* is lower for both, the older and the female group compared to the younger and male groups (see Figure 3a), which is in line with expectations based on [20][21]. Interestingly, older men and younger women report a higher technological anxiety (Figure 3b) whereas older subjects report of being in need of more professional training with ICT (Figure 4).

These significant differences give information about attitude towards technology in general and may also be used to



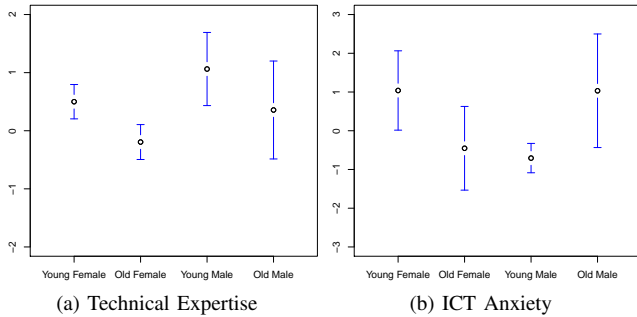


Figure 3. Self-reported Techn. Expertise (a) and ICT Riskiness (b).

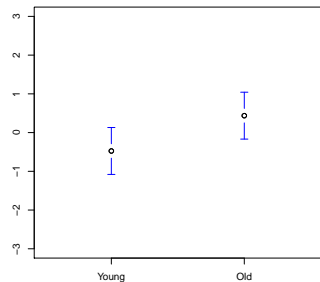


Figure 4. Self-reported Need for Training.

Table III  
SIGNIFICANT RESULTS FOR ASSESSMENTS OF THE MULTIMODAL SYSTEM. PEARSON'S R AND SIGNIFICANCE LEVEL (ASTERISK).

Data	ATT	PQ	HQI	HQS
Positive Technological Consequences (TA)	—	—	$r = .43^*$	—
Negative Technological Consequences (TA)	—	$r = -.44^*$	—	—

explain user group differences not easily explained by age or gender concerning interaction with the system, as well as differences in rating the system, which is analyzed in the following section (IV-A).

A. User group dependent rating of the multimodal system

The rating of the whole system was done after the last and most flexible interaction with the multimodal system. The four subscales of the AttrakDiff were used to assess the participants' evaluation of the whole system at that instance. The ratings of the subscales differs neither for age nor gender ( $\alpha = .05$ ).

Only when relying on the user group information, there are significant effects. Table III depicts the results of linear correlation analyses with the AttrakDiff subscales and the Part D user assessments as metrical variables:

- The Pragmatic Quality increases with lower expectations of Negative Consequences of technology.
- Hedonic Quality – Identity increases with assumed Positive technological Consequences.

Table IV  
SIGNIFICANT RESULTS FOR MODALITY PREFERENCES FOR AGE AND GENDER. F-VALUES ( $F_{(1,28)}$ ) AND SIGNIFICANCE LEVEL (ASTERISK).

Data	ATT	PQ	HQI	HQS	
Touch-Gesture	4.71*	—	—	—	gender
Voice-Gesture	12.27**	9.57**	6.84*	4.38*	
Touch-Voice	—	—	—	—	
Touch-Gesture	—	—	—	—	age
Voice-Gesture	—	—	—	—	
Touch-Voice	—	—	—	—	

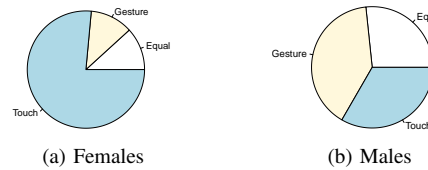


Figure 5. Modality preferences of ATT (Touch or Gesture).

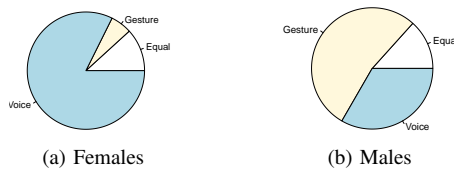


Figure 6. Modality preferences of ATT (Voice or Gesture).

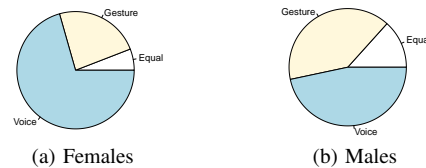


Figure 7. Modality preferences of PQ (Voice or Gesture).

B. User group dependent modality preferences

But what about modality preferences? Do the ratings of the single modality interactions (Part B) differ for the groups and user variables? This was tested for modality preferences as individual difference between the AttrakDiff subscales of all three modality pairs (Touch-Voice, Touch-Gesture, Voice-Gesture). Table IV summarizes the significant results of the ANOVAs; i.e., that gesture is preferred differently for gender, not for age. We decided to visualize the significant results categorically in Figures 5–9. It can be seen that the overall preference of using touch or voice over gestures concerning *Attractiveness* is dominant for female participants. A similar pattern is not as strong for the other three subscales. In contrast, male participants are divided concerning *Attractiveness*. Additionally, they prefer gestures over voice concerning *HQS*.

Can we get more insight into these results by analyzing

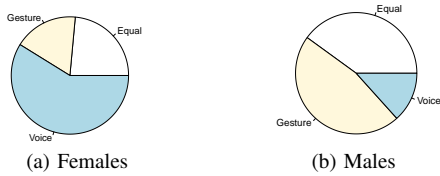


Figure 8. Modality preferences of HQI (Voice or Gesture).

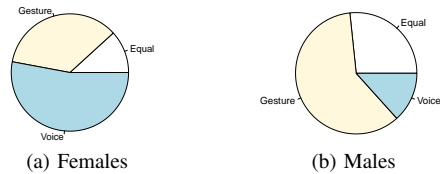


Figure 9. Modality preferences of HQS (Voice or Gesture).

Table V

SIGNIFICANT CORRELATIONS FOR MODALITY PREFERENCES WITH USER ASSESSMENTS. PEARSON'S R AND SIGNIFICANCE LEVEL (ASTERISK).

Data	Touch-Voice Technical Expertise	Voice-Gesture Technical Expertise
ATT	$r = .43^*$	$r = -.46^{**}$
PQ	$r = .51^{**}$	$r = -.56^{***}$
HQI	—	$r = -.38^*$
HQS	—	—

modality preferences with user variables assessed (see Table V)? The significant negative correlations between Technical Expertise (TA) and the preference of voice and touch over gesture (right column) are similar to the result of female participants preferring voice over gestures in general. Additionally, there is a significant correlation between Technical Expertise and preference of touch over voice (left column) not given for age nor gender. For touch and gesture there is no significant result. Also, neither the females' preference of voice over gesture on HQS and HQI with the opposite for males (Figures 8-9), nor the females' preference of touch over gesture on ATT (5) can be replicated with any of the user characteristics assessed by questionnaires. Thus, effects for Technical Expertise do not help to explain or further describe modality preference effects of gender. For example, the subscale Technical Expertise cannot significantly explain the inconclusive preferences between voice or touch and gesture for male participants (5, 6), although there is a visible tendency to prefer gesture with higher self-reported Expertise (see Figure 10).

V. SUMMARY AND DISCUSSION

When analyzing rating results of participants interacting with a smart-home system, the multimodal system was not judged differently for groups of age or gender. However, using questionnaire-based user characteristics, pragmatic quality increases with participants' decreasing assumed Negative

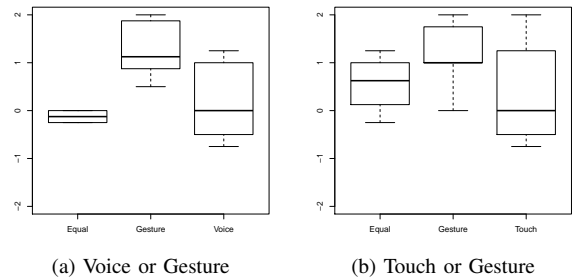


Figure 10. Boxplots of Technical Expertise and preferred modality (men).

Consequences of technology, and Hedonic Quality *Identity* correlates positively with assumed Positive Consequences of technology. Both user characteristics are not affected by age or gender, but give additional information on why individuals considered the interactive system as usable or to identify with it. The latter aspect may be considered even more relevant from a business point of view.

Regarding modality preferences based on rating differences of the unimodal interactions, there are significant results for gender and self-reported Technical Expertise: Whereas in general touch and voice are preferred over gestures, this result is valid only for female participants. Male subjects do not show clear preferences for touch or gesture (on the subscale ATT), and voice or gesture (ATT, PQ, HQS). On the two hedonic subscales, males even mostly prefer gesture over voice in opposition to females. Roughly comparable results are obtained with the self-reported Expertise information instead of gender.

However, as Technical Expertise is interrelated with age and gender, a final conclusion, which factor is causing the effect described – gender or Technical Expertise –, cannot be drawn. As the tendency of males' preference being dependent on Technical Expertise is not significant, a relevant effect size is not expected even with more subjects.

Not being directly related with the user group dependent ratings of the system and single modalities, results from the user assessments give rise to the question, why young female and old male participants seem to avoid ICT (Anxiety, Figure 3b) significantly more than old female and young male subjects, although this result is not in concordance with, e.g., Technical Expertise (Figure 3a).

VI. CONCLUSION AND FUTURE WORK

Even quite strong differences in age of adults do not result in different rating behavior of the multimodal interactive system or the preferences of single input modalities. Instead and surprisingly, gender seems to be a strong factor affecting modality preferences of unimodal interaction. These findings are opposed to results found in [6], which found limited influence of gender, especially considering age, however for task efficiency. In [3], there was no age effect for ratings after interaction, but a positive effect for females concerning

functional and usability aspects of the multimodal system used. Although limited in number of participants and limited to one experimental system (like [3]), the conclusion is to take gender into account much more for interactive systems than is done currently (e.g., [1][7][22]), especially when deciding on the investment into voice and/or gesture control. In this light, results of [22] that 55% of their subjects preferred controlling the home entertainment system via a GUI, but that users also stated that speech input would be their first choice if the speech recognizer had a lower error rate, would be interesting to reanalyze taking into account gender as well. Still, the nature of the interrelation between age, gender and technological expertise is still to be identified with, e.g., special recruited participants. From a pragmatic point of view, grouping users according to gender is much easier than assessing technical expertise.

Furthermore, using other assessment methods will be necessary for answering the questions raised here. For example, addressing the impact of degrees of cognitive abilities was not possible with our recruited participants, although beneficial to the purpose of this paper. Also, we observed single older adults having trouble using the touch screen efficiently. This did not affect the results, but for additional experiments with older adults, assessing dexterity, e.g., with the Grooved Pegboard seems to be advisable.

#### REFERENCES

- [1] M. Turunen, A. Melto, J. Hella, T. Heimonen, J. Hakulinen, E. Mäkinen, T. Laivo, and H. Soronen, "User expectations and user experience with different modalities in a mobile phone controlled home entertainment system," in *MobileHCI*, 2009.
- [2] M. Turunen and T. Hakulinen, J. Heimonen, "Assessment of spoken and multimodal applications: Lessons learned from laboratory and field studies," in *Interspeech*, 2010, pp. 1333–1336.
- [3] T. Jokinen, K. und Hurtig, "User expectations and real experience on a multimodal interactive," in *Interspeech*, 2006, pp. 1049–1052.
- [4] A. Naumann, I. Wechsung, and S. Möller, "Factors influencing modality choice in multimodal applications," in *Perception in Multimodal Dialogue Systems*, 2008, pp. 37–43.
- [5] D. Petrelli, A. De Angeli, W. Gerbino, and G. Cassano, "Referring in multimodal systems: the importance of user expertise and system features," in *ACL Workshop on Referring Phenomena*, 1997, pp. 14–19.
- [6] A. Naumann, I. Wechsung, and J. Hurtienne, "Multimodal interaction: A suitable strategy for including older users?" *Interacting with Computers*, vol. 22, no. 6, pp. 465–474, 2010.
- [7] K. Jeong, R. Proctor, and G. Salvendy, "A survey of smart home interface preferences for U.S. and Korean users," in *Human Factors and Ergonomics Society Annual Meeting*, vol. 53, 2009, pp. 541–545.
- [8] T. Kleinberger, M. Becker, E. Ras, A. Holzinger, and P. Müller, "Ambient intelligence in assisted living: Enable elderly people to handle future interfaces," in *Universal Access in HCI, HCI International*, 2007, pp. 103–112.
- [9] M. Perry, A. Dowdall, L. Lines, and K. Hone, "Multimodal and ubiquitous computing systems: Supporting independent-living older users information technology in biomedicine," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, pp. 258–270, 2004.
- [10] M. Zajicek and W. Morrissey, "Multimodality and interactional differences in older adults," *Universal Access in the Information Society*, vol. 2, pp. 125–133, 2003.
- [11] W. Rogers and A. Fisk, "Toward a psychological science of advanced technology design for older adult," *Journals of Gerontology*, vol. 65B, no. 6, pp. 645–653, 2010.
- [12] A. Murata and H. Iwase, "Usability of touch-panel interfaces for older adults," *Journal of the Human Factors and Ergonomics Society*, vol. 47, no. 4, pp. 767–776, 2005.
- [13] J. Nielsen, *Usability Engineering*. San Francisco: Morgan Kaufmann, 1993.
- [14] C. Kühnel, B. Weiss, and S. Möller, "Evaluating multimodal systems – A comparison of established questionnaires and interaction parameters," in *NordiCHI*, 2010, pp. 286–293.
- [15] N. Tractinsky, A. Cokhavi, M. Kirschenbaum, and T. Sharfi, "Evaluating the consistency of immediate aesthetic perceptions of web pages," *International Journal on Human-Computer Studies*, vol. 64, pp. 1071–1083, 2006.
- [16] D. Wechsler, *Manual for Wechsler Memory Scaled – Revised*. New York: Psychological Corporation, 1981.
- [17] K. Karrer, C. Glaser, C. Clemens, and C. Bruder, "Technikaffinität erfassen – der Fragebogen TA-EG," in *Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme*, ser. ZMMS Spektrum, vol. 22, no. 29. Düsseldorf: VDI Verlag, 2009, pp. 196–201.
- [18] M. Hassenzahl and A. Monk, "The inference of perceived usability from beauty," *Human-Computer Interaction*, vol. 25, no. 3, pp. 235–260, 2010.
- [19] M. Hassenzahl, "The interplay of beauty, goodness, and usability in interactive products," *Human-Computer Interaction*, vol. 19, pp. 319–349, 2008.
- [20] J. Cooper and M. Kugler, "The digital divide: The role of gender in human-computer interaction," in *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, 2nd ed. New York: Lawrence Erlbaum, 2008, pp. 763–775.
- [21] S. Czaja and C. Lee, "Information technology and older adults," in *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, 2nd ed. New York: Lawrence Erlbaum, 2008, pp. 777–792.
- [22] M. Johnston, L. D Haro, M. Levine, and B. Renger, "A multimodal interface for access to content in the home," in *Annual Meeting of the Association for Computational Linguistics*, 2007, pp. 376–383.

## Users' Trust and Secure Feeling towards Cloud Services

Kaarina Karppinen

Software Technologies, Security Assurance  
 VTT Technical Research Centre of Finland  
 Oulu, Finland  
 kaarina.karppinen@vtt.fi

**Abstract**—The paper describes the study in which end users' trust and feeling of security in regard to cloud services are examined. Trust and security are one of the key issues within cloud where people use a shared pool of computing resources for the storage of their data and thus do not have control over the management of their own data content. An online survey was conducted in three countries (Finland, USA and Japan) with more than 3000 respondents. The twelve cloud trust creators that were identified by cloud software experts in our earlier study were picked, and asked the end users to rank how important those factors are for them when deciding if it is secure to use a service in the cloud. The results show that ease of use has the biggest effect on the secure feeling towards cloud services, followed by the language and the price of the service. Some differences were visible between the countries. These as well as the differences for the results between genders, age groups and expertise levels of the respondents are presented in this paper.

*Keywords*-cloud services; end users; security; trust

### I. INTRODUCTION

Within the cloud, software, infrastructure and platforms are offered as a service. [1] In cloud services, content (for example, music, email, files), personal information and programs are stored on Internet servers instead of the user's own computer. [2] The cloud services and the content in them are accessible from various devices anytime and anywhere through the Internet.

The cloud environment relies heavily to the notion of trust. People often find it harder to trust on-line services than off-line services. [3] Cloud computing introduces a whole ecosystem of clients, services and infrastructure, where trust boundaries are moved into realms where physical locations and even ownerships are unknown. [4] Establishing, expressing and maintaining trust by technological methods has always been challenging. The traditional question on where are the limits to presenting, simplifying and visualising security without losing trust or creditability, receive a new life within the cloud environment. [5]

Since there are no deterministic means of revealing the factors affecting the human perception of security or trust on cloud environment, security experts and user experience experts who have some familiarity with clouds were first interviewed. Their views on the role of trust in cloud services were then analysed. After that, the results of these 33 interviews [5] were verified by conducting an online survey

with end users. Our objective for this study was to investigate what aspects and to what extent contribute and affect to the secure feeling in the cloud.

The rest of this paper is organised as follows. In Section 2, the background literature in relation to cloud services as well as trust is discussed. Section 3 describes the research methodology used in this study. Section 4 presents study results which are then further discussed in Section 5. Section 6 summarizes the study results.

### II. BACKGROUND

In this section, a brief background to two main subjects of the study – cloud services and trust – is presented.

#### A. Cloud Services

Cloud services and cloud computing are today's reality; typical web users use cloud services daily whether they know it or not. [2] The cloud concept is still changing but several early definitions for cloud computing and cloud services exist.

Weiss [6] states that cloud computing is a paradigm where software functionality, hardware computing power, and other computing resources are delivered in the form of service so that they become available widely to consumers. In [7], the following definition is constructed: Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically re-configured to adjust to a variable load (scale), allowing also for an optimum resource utilization. Breiter and Behrendt [8] take the user point of view to cloud services: "The emerging style of cloud computing provides applications, data, and information technology resources as services of a network. The cloud services approach focuses on a positive user experience while shielding the user from the complexity of the underlying technology." In [9], it is continued that services such as free email services, Internet portal services, web hosting services, computing infrastructure services, etc. are all cloud services. NIST [10] defines cloud services from the consumers' perspective as follows: "The capability provided to the consumer to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems,

storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.”

The main message for the cloud users based on various cloud definitions seems to be that, in the cloud, the end users do not need to worry about the underlying technology and infrastructure details but just rely on the cloud services as a huge storage area of their data, resources and software. This relying is calling for trust.

### B. Trust and Secure Feeling

Although a concise and universally accepted definition of trust has remained elusive [11][12], most definitions include an aspect of perceived risk of vulnerability, and also an element of psychological security.

Trust can be said to be a belief, attitude, or expectation concerning the likelihood that the actions or outcomes of another individual, group or organization will be acceptable or will serve the actors interests. [13]

In [14], it is said that trust is letting other persons (natural or artificial, such as firms, nations, etc.) take care of something the trustor cares about, where such 'caring for' involves some exercise of discretionary powers.

Fogg and Tseng [15] define trust as an indicator of a positive belief about the perceived reliability of, dependability of, and user confidence in himself being able to control the technology.

Trust assessment on the cloud environment is complicated by the fact that in the digital world one is faced with the absence of physical cues and the establishment of suitable centralized authorities such as certification authorities is still evolving and not always applicable. [16]

In [17], it is summarised that trust becomes increasingly important as mobile services get more and more involved in the personal life of people and as these services increasingly collect, analyse and store personal data. This is just as well true with not only mobile but all cloud based services.

Building trust in the cloud is not a straightforward task and there are a lot of issues to be solved before a user can truly feel safe towards services in the cloud.

## III. RESEARCH METHOD

As the target of the study was to examine the feelings of cloud service end users, and as the cloud is a worldwide concept, a broad and representative sample of end users in several countries was needed. Hence, a quantitative research approach was chosen. A web-based survey was used to conduct the quantitative study. According to [18] quantitative data enable standardized and objective comparisons and the measurement of quantitative research permits overall descriptions of situations or phenomenon in a systematic and comparable way. There are several advantages of web-based surveys. As stated in [19], in a web-based survey the speed of recruitment is very fast and travel cost and time are eliminated. [19] This was important as an international study with respondents in several different countries was meant to be conducted.

A web-based survey was seen as an appropriate way to go deeper from the interview results gotten earlier. [5] It

must be noted that the survey consisted of several other dimensions beside the trust and security issues; e.g., our results of today's usage of cloud services based on the survey are already published in [2].

### A. Data Sources and Collection

The survey design process adopted from [19] and [20] was followed. When designing the sampling for the survey the sampling design process adopted from [20] was used.

The survey was decided to be conducted in three different countries: Finland, Japan and USA. Decision of choosing these three countries was due to the fact that each of them had acted as a pacesetter in many ICT fields, e.g., in mobile technology and services usage. As these countries have been in the forefront of technology use and acceptance in the consumer sector, they present fruitful contexts for analyzing consumers' use of and perceptions towards new technologies such as cloud services. Furthermore, all of these three countries are so dissimilar with each other concerning the location and the cultural background that it was thought to be interesting to find out possible differences when comparing the results of these countries.

The original survey was made in English; the Finnish and Japanese versions were translated by experts. The questionnaires for each of the three countries were identical except for the background question about the annual income of the respondents where the currency was localised for each country. Each of the three language versions was then created to be an individual online survey by one market research company in each of the countries who also took care of finding the respondents matching the pre-defined criteria from their customer registers and closing the polls after the desired sample size was acquired.

The sample size was decided to be approximately one thousand respondents from each country. This size is typical for surveys that are nationwide, regardless of the size of the country. Thus, the overall sample size was ca 3000 responses. The survey was generated and data gathered with Digium Enterprise data collection service. A total of 3187 responses, all of them acceptable, was received between the time period of 10 June and 21 June 2010. There were 1005 respondents from Finland, 1089 from USA and 1093 from Japan.

The criterion for selecting the survey respondents was to get a representative sample from each country based on the demographics of the countries. The gender, age and self evaluated cloud service expertise distribution of respondents from different countries can be seen in Table 1. Furthermore, the demographic information of the gender and age of the population for each country is shown in brackets (for more information see [21][22][23]).

### B. Data Analysis Method

The quantitative analysis of the survey was made using SPSS Statistics 17.0 computer software. At the first stage of the analysis frequency distributions and other descriptive statistics were obtained. Also several cross tabulations, t-tests and analysis of variances (ANOVA) were made to compare the samples.

TABLE I. GENDER, AGE AND EXPERTISE DISTRIBUTION OF THE RESPONDENTS

n	Finland	USA	Japan
	1005	1089	1093
<b>Gender</b>	%	%	%
Male	50,3 (48,6)	47,7 (49,1)	50,4 (49,8)
Female	49,7 (51,4)	52,3 (50,9)	49,6 (50,2)
<b>Age</b>	%	%	%
15-24	11,2 (14,8)	13,3 (17,7)	13,7 (15,7)
25-34	19,6 (15,3)	16,5 (18,0)	17,7 (16,7)
35-44	15,8 (15,0)	19,9 (20,4)	21,1 (18,1)
45-54	20,6 (16,9)	19,2 (17,0)	16,7 (15,6)
55-64	20,4 (17,6)	15,5 (11,0)	21,0 (18,9)
Over 65	12,3 (20,4)	15,5 (15,8)	9,8 (15,0)
<b>Expertise</b>	%	%	%
Novice user	60,7	70,6	83,0
Intermediate user	32,3	25,3	15,9
Advanced user	7,0	4,1	1,1

IV. SECURE FEELING IN THE CLOUD

Our research question regarding trust and secure feeling in the cloud was about the factors that could contribute and affect to the user’s secure feeling in the cloud. The question was phrased in the following way:

*How important are the following factors when deciding if it is secure to use a specific cloud service (for example, webmail or music library)?*

A 7 step Likert scale, where 1=Not important at all, and 7=Extremely important, was used. Also a Not applicable option was given. The cloud trust affect factors found in our earlier interview [5] were used as a basis for the options which the respondents was asked to value. These 12 affect factors were:

- Own experiences
- Friends’ recommendation
- Brand reputation, image and name
- Background and home country of the company
- A critical mass of users
- Search engine results
- User license agreements
- Price
- Ease of use
- Visual image of the service web page
- Transparency
- Language

A. Overall Results

According to the survey, the factors having an effect on the overall users’ experience about the secure usage of cloud services are ranked in the order of importance in Table 2.

When looking at the average results of the whole respondent group the most significant factor in trust creation is stated to be the ease of use. It received the mean value of 5,91, Two out of every five (40,1 %) of the respondents had valued this factor as Extremely important (value 7 in the Likert scale).

The ease of use was closely followed by the language issues. The respondents feel that when deciding if it is secure

to use a specific cloud service the fact that the service uses their own language is important by an average of 5,86. As many as 43,1% of the respondents had given the highest importance value for this factor.

Maybe the most surprising factor is the price, which was ranked as the third in the most important factor list ranked by the average values, with the mean of 5,72, and 37,1% of the respondents given the highest possible importance value to this factor.

When taking a look at the last factors on that list it is noted, also quite surprisingly, that the recommendation from one’s friends receives the lowest markings out of these suggested factors.

B. Results divided by countries

These twelve affect factors divided by countries are presented in Figure 1. The ease of use as well as understandable user license agreements are appreciated the most in Finland. The language issues are considered extremely important, especially in USA. For the American respondents it is also significantly more important that the company is located in their home country than for the Finnish or Japanese respondents. Finnish respondents value their own and their friends’ experiences more than other two countries; actually, in Japan, the recommendation from one’s friends receives the lowest markings out of all the questions and all the countries. The Japanese respondents, on the other hand, have more faith in the search engine results than the other two countries, and do not care as much about the appearance of the web page as the others.

TABLE II. THE FACTORS OF SECURE FEELING WITHIN CLOUD SERVICES

How important are the following factors when deciding if it is secure to use a specific cloud service:	Mean	Median	Std. Deviation
	The service is easy to use.	5,91	6,00
The service uses my own language.	5,86	6,00	1,521
The price is low.	5,72	6,00	1,528
The user license agreements are understandable.	5,54	6,00	1,522
The appearance of the service web page is nice and professional.	5,32	6,00	1,475
The company offering the service has a good brand reputation, image and name.	4,84	5,00	1,621
I have previous good experiences about the service.	4,70	5,00	1,650
The company offering the service is located in my home country.	4,60	5,00	1,829
Many of the people I know use the service.	4,49	5,00	1,708
I am able to look through the source code of the service.	4,37	4,00	1,915
The service is one of the first hits within the search engine results.	4,19	4,00	1,755
My friends recommend the service.	4,08	4,00	1,804

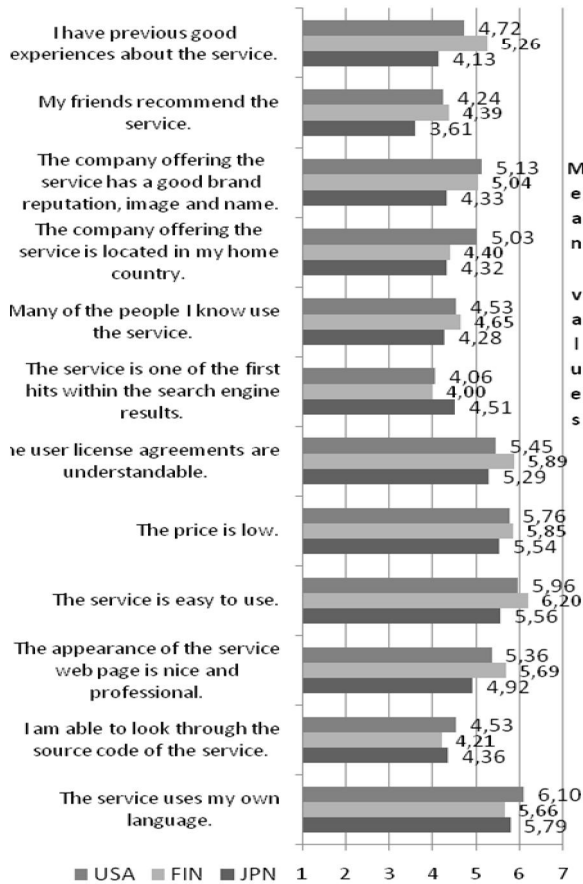


Figure 1. The factors of secure feeling within cloud services divided by countries.

C. Results divided by genders, age groups and expertise levels of the respondents

As can be seen in Figure 2, the main difference between the genders was that the female respondents had given higher values to all the factors than the males.

No similar dependencies could be seen when comparing the results between the age groups, presented in Table 3. However, there were significant differences in some of the factors, especially between the youngest and oldest age groups. When looking at the most evident (more than half a unit difference between the averages) differences between the age groups it can be noticed that the youngsters and the elderly people have different attitudes towards a few of the factors. The young people (below 35 years) respect their own and their friends experiences more than the old ones (over 65 years) while the opposite is true with the home country and language of the service. The results between the different self-evaluated expertise groups are shown in Figure 3.

For the novice users, their own and their friends' experiences, appearance of the service web page, or company's reputation were not as important as for the more advanced users when deciding if it secure to use a specific

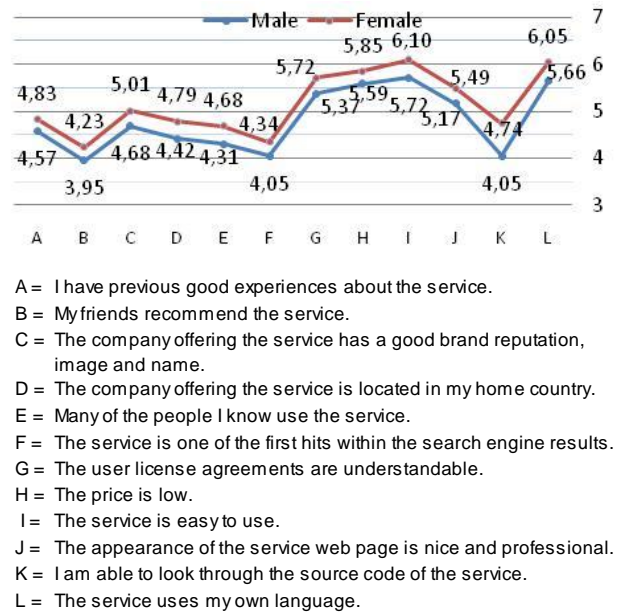


Figure 2. The factors of secure feeling within cloud services divided by genders.

cloud service. On the other hand, the language of the service or the location of the service provider did receive lower average values from the advanced users than from the novice or intermediate users. These results imply that users prefer to begin their trust creation process by gaining experiences from local service providers who offer services in users' mother tongue. Only when enough experience has been collected can the users have confidence to their capabilities to judge the security of a service based on their prior experiences and knowledge.

The results divided between the age groups and the expertise are strikingly similar. This of course is partly due to the fact that there is a correlation between those to variables; a bigger percentage of the older people judged themselves to be novice users than of the younger people.

V. DISCUSSION

People are already in the cloud even if they might not know it. They use cloud services such as webmail or social networking sites without thinking of the underlying technology. Majority of cloud end users still evaluate themselves as novice users. [2] In this environment people are drawn into services that they trust and feel secure towards.

Our objective for this study was to investigate what aspects and to what extent contribute and affect to the secure feeling in the cloud.

The fact that ease of use got the highest average value as a factor affecting to end user's trust and secure feeling towards services in the cloud is by no means an unexpected result. Already, Fogg and Tseng [15] noted that one of the elements included in user trust experience is that the

TABLE III. FEELING OF SECURITY FACTORS WITHIN CLOUD SERVICES DIVIDED BY AGE GROUPS

How important are the following factors when deciding if it is secure to use a specific cloud service:						
Age:	15-24	25-34	35-44	45-54	55-64	65+
I have previous good experiences about the service.	5,04	4,91	4,65	4,69	4,47	4,36
My friends recommend the service.	4,51	4,51	4,13	3,91	3,70	3,67
The company offering the service has a good brand reputation, image and name.	5,00	4,80	4,75	4,90	4,82	4,84
The company offering the service is located in my home country.	4,17	4,27	4,53	4,71	4,89	5,14
Many of the people I know use the service.	4,78	4,72	4,43	4,37	4,41	4,18
The service is one of the first hits within the search engine results.	4,50	4,26	4,04	4,06	4,24	4,08
The user license agreements are understandable.	5,35	5,43	5,42	5,63	5,72	5,70
The price is low.	5,76	5,67	5,68	5,75	5,77	5,67
The service is easy to use.	5,84	5,83	5,82	6,00	6,00	5,95
The appearance of the service web page is nice and professional.	5,36	5,35	5,25	5,41	5,26	5,34
I am able to look through the source code of the service.	4,26	4,18	4,27	4,51	4,49	4,60
The service uses my own language.	5,60	5,50	5,88	5,98	6,08	6,09

users can trust themselves of being able to control the technology. Also, Kaasinen [17] states that the users have to have a clear conception of the functionality of the service even if they do not need to know all the details. In [24], convenience of use was identified as one of two key success factors of mobile Internet services.

It is interesting, though, that ease of use showed to be of different importance level with different respondent groups. This factor has significantly more importance for the Finns than to the Japanese respondents. Some of this may be explained by the fact that this factor was more important for the advanced users of cloud services than for the novice users, and in Finland a considerably larger percentage of respondents evaluated themselves to be advanced cloud service users than in Japan.



Figure 3. The factors of secure feeling divided by end users' self-evaluated experience within cloud services.

The ease of use was closely followed by the language of the service. Especially the American respondents felt strongly that when deciding if it is secure to use a specific cloud service it is important that the service uses their own language. This is quite obvious as English language has such a strong ground in ICT sector. The end users have grown used to using English in the internet environment.

It was somewhat surprising that the low price was ranked to be the third important factor in this context. Conventionally, low price is thought not to increase trust. It might be that the fact that many of the largely popular cloud services (e.g., Facebook, Gmail and Youtube) are free has affected the general atmosphere and consumers are willing to lay trust on the free or low priced services more than before cloud era. However, low price has also earlier been noted to act as one of the key success factors of a specific internet solution. [25]

Limitations of the study mainly rise from the fact that our study was conducted as an online survey, which excludes the group of people not active in the Internet. Nevertheless, as the target of the survey was to study the end users' attitude towards online services it was only natural to choose the online method for conducting the survey.



VI. CONCLUSION AND FUTURE WORK

Our study examined end users' trust and secure feeling in regard to cloud services. An online survey was conducted in three countries (Finland, USA and Japan) with more than 3000 respondents. The twelve factors that create trust in the cloud, which were identified by cloud software experts in our earlier study, were taken and asked the end users to rank how important those factors are for them when deciding if it is secure to use a specific cloud service.

The results show that in general the ease of use has the biggest effect on the feeling of security towards cloud services, followed by the language and the price of the service.

Some differences were visible between the countries. The ease of use as well as understandable user license agreements are appreciated the most in Finland. The language issues are considered extremely important especially in USA. For the American respondents it is also significantly more important that the company is located in their home country than for the Finnish or Japanese respondents. The Japanese respondents, on the other hand, have more faith in the search engine results than the other two countries, and do not care as much about the appearance of the web page as the others.

The main difference between the genders was that the female respondents had given higher values to all the factors than the males. The differences between the age groups and expertise were significant between the youngest and oldest respondents as well as the self-evaluated novice and advanced cloud service users. The young as well as the advanced users respect their own and their friends' experiences more than the oldest age group and the novice users, while the opposite is true with the home country and language of the service.

These results will be taken a deeper look at our future research where we will be studying end users' perceptions towards cloud services and their security using an online focus group research method.

ACKNOWLEDGMENT

This study was carried out in Cloud Software Program launched by Finnish Strategic Centre for Science, Technology and Innovation TIVIT Plc and conducted together with F-Secure. We thank both of these parties for their support.

REFERENCES

[1] B. Hayes, "Cloud computing," in *Communications of the ACM*, 2008, vol. 51.

[2] K. Karppinen, K. Koskela, and T. Koivumäki, "Usage of Cloud Services - A Survey for Consumers in USA, Japan and Finland". 1st International Conference on Cloud Computing and Services Science - CLOSER2011, Noordwijkerhout, The Netherlands, May 7 - 9, 2011, pp. 305-308.

[3] S. J. Best, B. S. Kreuger, and J. Ladewig, "The effect of risk perceptions on online political participatory decisions." *Journal of Information Technology & Politics*, vol. 4, 2005.

[4] TiViT, "Cloud software program description," <http://www.cloudsoftwareprogram.org/cloud-security>. Accessed 20 November 2011

[5] I. Uusitalo, K. Karppinen, A. Juhola, and R. Savola, "Trust and cloud services - an interview study". 2nd IEEE International Conference on Cloud Computing Technology and Science, CloudCom2010, Indianapolis, IN, USA 30 November–3 December 2010, pp. 712-720.

[6] A. Weiss, "Computing in the Cloud", *netWorker*, Vol. 11, No. 4, pp. 16-26, 2007.

[7] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition." *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 1, Dec 2008.

[8] G. Breiter and M. Behrendt, "Life cycle and characteristics of services in the world of cloud computing", *IBM Journal of Research and Development*, v. 53 n. 4, pp. 527-534, July 2009

[9] W. Kim, S.D. Kim, E. Lee, and S. Lee, "Adoption issues for cloud computing". 11th International Conference on Information Integration and Web-based Applications & Services, pp. 3-6. doi:10.1145/1806338.1806341

[10] National Institute of Standards and Technology, Information Technology Laboratory, The NIST Definition of Cloud Computing, Retrieved April 18, 2010, from <http://csrc.nist.gov/groups/SNS/cloud-computing/>

[11] W.E.D. Creed and R.E. Miles, "Trust in organizations: a conceptual framework linking organizational forms, managerial philosophies, and the opportunity costs of controls". In: Kramer, R.M., Tyler, T.R. (Eds.), *Trust in Organizations: Frontiers of Theory and Research*. 1995, Sage Publications, London, pp. 16–38.

[12] R.M. Kramer and K.S. Cook, *Trust and Distrust in Organizations: Dilemmas and Approaches*, Russell Sage Foundation, New York, 2004.

[13] S. Sitkin and N. Roth, "Explaining the limited effectiveness of legalistic 'remedies' for trust/distrust." *Organization Science*, 1993.

[14] A. Baier, "Trust and antitrust," *Ethics*, 1986.

[15] B.J. Fogg and H. Tseng, "The elements of computer credibility". CHI 99 Conference, Pittsburgh, Pennsylvania, USA, 15 – 20 May 1999, ACM Press, pp. 80 – 87.

[16] E. Chang, T. Dillon, and D. Calder, "Human system interaction with confident computing. the mega trend." *Conference on Human System Interactions*, 2008.

[17] E. Kaasinen, "User Acceptance of Mobile Services – Value, Ease of Use, Trust and Ease of Adoption," doctoral dissertation, VTT, Espoo, Finland, 2005, VTT Publications 566, <http://www.vtt.fi/inf/pdf/publications/2005/P566.pdf>.

[18] K.F. Punch, *Introduction to Social Research – Quantitative and Qualitative approaches*. SAGE Publications, Great Britain, 1998.

[19] V. Kumar, D.A. Aaker, and G.S. Day. *Essentials of Marketing Research*. Donnelley Willard, United States, 2nd edition, 2001.

[20] N.K. Malhotra, *Marketing research and applied orientation*. Prentice-Hall, Inc., United States, 2nd edition, 1996.

[21] Statistical Survey Department, Statistics Bureau, Ministry of Internal Affairs and Communications, Japan, 2010. *Population of Japan - Final Report of the 2005 Population Census*. Available at: <http://www.stat.go.jp/data/nenkan/pdf/yhyou02.pdf> Accessed 20 November 2011

[22] Statistics Finland, 2010. *Population by the age in the end of 2009*. Available at: [http://www.stat.fi/tup/suoluk/suoluk\\_vaesto.html](http://www.stat.fi/tup/suoluk/suoluk_vaesto.html) Accessed 20 November 2011

[23] U.S. Department of Commerce, 2010. *Profiles of General Demographic Characteristics; 2000 Census of Population and Housing*. Available at: <http://www.census.gov/census2000/demoprofiles.html> Accessed 20 November 2011

[24] T.S.H. Teo and S.H. Pok, "Adoption of the Internet and WAP-enabled phones in Singapore". *Behaviour and Information Technology*, 2003, 22, pp. 281 – 289.

[25] S.J. Barnes and S.L. Huff, "Rising sun: iMode and the wireless Internet", *Communications of the ACM*, Vol. 46, no. 11, 2003, pp. 78 – 84.